Utilizing self-similar stochastic processes to model rare events in finance

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)
eingereicht an der
Wirtschaftswissenschaftlichen Fakultät der
Humboldt-Universität zu Berlin

Niels Wesselhöfft

Präsidentin: Dr.-Ing. Dr. Sabine Kunst
Dekan: Dr. Daniel Klapper
Erster Gutachter: Dr. Wolfgang Karl Härdle
Zweiter Gutachter: Dr. Weining Wang
Tag des Kolloquiums: 14. Februar 2020

Abstract:

Coming from a sphere in statistics and mathematics in which the Normal distribution is the dominating underlying stochastic term for the majority of the models, we indicate that the relevant diffusion, the Brownian Motion, is not accounting for three crucial empirical observations for financial data: Heavy tails, long memory and scaling laws. A self-similar process, which is able to account for long-memory behavior is the Fractional Brownian Motion, which has a possible non-Gaussian limit under convolution of the increments. The increments of the Fractional Brownian Motion can exhibit long memory through a parameter H, the Hurst exponent. For the Fractional Brownian Motion this scaling (Hurst) exponent would be constant over different orders of moments, being unifractal. But empirically, we observe varying Hölder exponents, the continuum of Hurst exponents, which implies multifractal behavior. We explain the multifractal behavior through the changing α-stable indices from the α-stable distributions over sampling frequencies by applying filters for seasonality and time dependence (long memory) over different sampling frequencies, starting at high-frequencies up to one minute. By utilizing a filter for long memory we show, that the low-sampling frequency process, not containing the time dependence component, can be governed by the α-stable motion. Under the α-stable motion we propose a semiparametric method coined Frequency Rescaling Methodology (FRM), which allows to rescale the filtered high-frequency data set to the lower sampling frequency. The data sets for e.g. weekly data which we obtain by rescaling high-frequency data with the Frequency Rescaling Method (FRM) are more heavy tailed than we observe empirically. We show that using a subset of the whole data set suffices for the FRM to obtain a better forecast in terms of risk for the whole data set. Specifically, the FRM would have been able to account for tail events of the financial crisis 2008.

Abstract: (German)

In der Statistik und der Mathematik ist die Normalverteilung der am meisten verbreitete, stochastische Term für die Mehrheit der statistischen Modelle. Wir zeigen, dass der entsprechende stochastische Prozess, die Brownsche Bewegung, drei entscheidende empirische Beobachtungen nicht abbildet: schwere Ränder, Langzeitabhängigkeiten und Skalierungsgesetze. Ein selbstähnlicher Prozess, der in der Lage ist Langzeitabhängigkeiten zu modellieren, ist die Gebrochene Brownsche Bewegung, welche durch die Faltung der Inkremente im Limit nicht normalverteilt sein muss. Die Inkremente der Gebrochenen Brownschen Bewegung können durch einen Parameter H, dem Hurst Exponenten, Langzeitabhängigkeiten darstellt werden. Für die Gebrochene Brownsche Bewegung müssten die Skalierungs- (Hurst-) Exponenten über die Momente verschiedener Ordnung konstant sein. Empirisch beobachten wir variierende Hölder-Exponenten, die multifraktales Verhalten implizieren. Wir erklären dieses multifraktale Verhalten durch die Änderung des α-stabilen Indizes der α-stabilen Verteilung, indem wir Filter für Saisonalitäten und Langzeitabhängigkeiten über verschiedene Zeitfrequenzen anwenden, startend bei 1-minütigen Hochfrequenzdaten. Durch die Anwendung eines Filters für die Langzeitabhängigkeit zeigen wir, dass die Residuen des stochastischen Prozesses geringer Zeitfrequenz (wöchentlich) durch die α-stabile Bewegung beschrieben werden können. Dies erlaubt es uns, den empirischen, hochfrequenten Datensatz auf die niederfrequente Zeitfrequenz zu skalieren. Die generierten wöchentlichen Daten aus der Frequenz-Reskalierungs-Methode (FRM) haben schwerere Ränder als der ursprüngliche, wöchentliche Prozess. Wir zeigen, dass eine Teilmenge des Datensatzes genügt, um aus Risikosicht bessere Vorhersagen für den gesamten Datensatz zu erzielen. Im Besonderen wäre die Frequenz-Reskalierungs-Methode (FRM) in der Lage gewesen, die seltenen Events der Finanzkrise 2008 zu modellieren.

Contents

Chapter 1

Introduction

1.1 Scope

There appears to be a severe misunderstanding in modern statistics. A statistical or econometric model aims to account for systematic features in the data, reflecting a subset of the data generating process. Consequently, if the feature is well modelled, the according model errors, the residuals, should not contain this feature anymore. If this characteristic is presumed to be prevalent over time and scale, a good model should be able to account for the characteristic, not only over subsets, but also over scale, i.e. sampling frequencies. Consequently the underlying parameters of the model should converge to a constant as more data are available. The model will be robust, if the underlying parameters are not significantly varying over subsets and sampling frequencies. This certainly includes starting values and hyperparameters of potentially highly nonlinear models like Neural Networks or Support Vector Machines. Wandering into the field of non-stationary models, accepting or trying to explain dynamics of parameters itself appears to be a distortion of the original model goal. The consequence is going to be that the model is a higher order function approximation of the feature, which will, with large

probability, not be robust afterall. Instead the thesis will focus on self-similar structures, especially stochastic processes, which have well defined properties, accounting for heavy tails and time dependence over sampling frequencies resting on generalizations of the central limit theorem in the α-stable framework. The relevant models are applied to portfolio theory and risk management in the domain of finance, which involves modeling especially high-frequency stock returns in order to estimated the horizon distribution, such as weekly, monthly and yearly.

1.2 Power laws and self-similarity

The power law states that the frequency of an event given its size $f(x)$ is inversely proportional to the power d of the size (Clauset et al., 2009).

$$f(x) = x^{-d} \tag{1.1}$$

It follows that scaling the size with a factor c, which might be length, energy or time, implies a proportionate scaling of the frequency $f(x)$.

$$f(cx) = (cx)^{-d} = c^{-d} f(x) \propto f(x) \tag{1.2}$$

This property of power laws is called scale invariance.[1] The power law can be linearized by taking logarithms,

$$\log \{f(x)\} = -d \log(x) \tag{1.3}$$

where d denotes the fractal dimension.

Geometry

In geometry an object K is called self-similar if the object looks roughly to be the same over different scales. An object K consists of a union of $m \in \mathbb{N}^+$ similar

[1] $f(cx) \propto f(x)$ or $f(cx) \sim f(x)$ indicates direct proportionality.

objects $f(K)$, where f is a function called similarity.

$$K = \bigcup_{i=1}^{m} f_i(K) \tag{1.4}$$

The according power law with the number of new objects $N \in \mathbb{N}^+$ and shrinking in size by factor $s \in \mathbb{R}$ provides

$$N = (1/s)^{-d} \tag{1.5}$$

the Hausdorff dimension $d \in \mathbb{R}$.

$$d = \lim_{1/s \to 0} \frac{\log N}{\log 1/s} \tag{1.6}$$

The square for example generates $N = 4$ new segments in the first iteration. The length of the object decreases by factor $s = 1/2$, so $1/s = 2$. The according Hausdorff dimension is $d = \frac{\log 4}{\log 2} = 2$. The Sierpinski triangle (see Figure 1.1) generates $N = 3$ new objects, $1/s = 2$, implying a fractal dimension of $d = \frac{\log 3}{\log 2} \approx 1.585$. Hence, the introduction of the Hausdorff dimension offers a wider field of geometric objects with appealing properties that are able to depict phenomena of different scientific fields.

Stochastic processes

For stochastic processes in economics the empirical scaling law states that the average absolute (squared) price changes, as functions of their time intervals, are proportional to a power of the interval size (Mandelbrot, 1982; Mantegna and Stanley, 1995; Mandelbrot et al., 1997b; Calvet and Fisher, 2002; Xu and Gencay, 2003). Starting from the self-affine process $\{X_t\}$, $t \geq 0$ with Hurst exponent $H > 0$ and $c > 0$,

$$X_{ct} \stackrel{\mathcal{L}}{=} c^H X_t \tag{1.7}$$

$$\mathsf{E}\{|X_{ct}|^p\} = c^{Hp}\,\mathsf{E}\{|X_t|^p\} \tag{1.8}$$

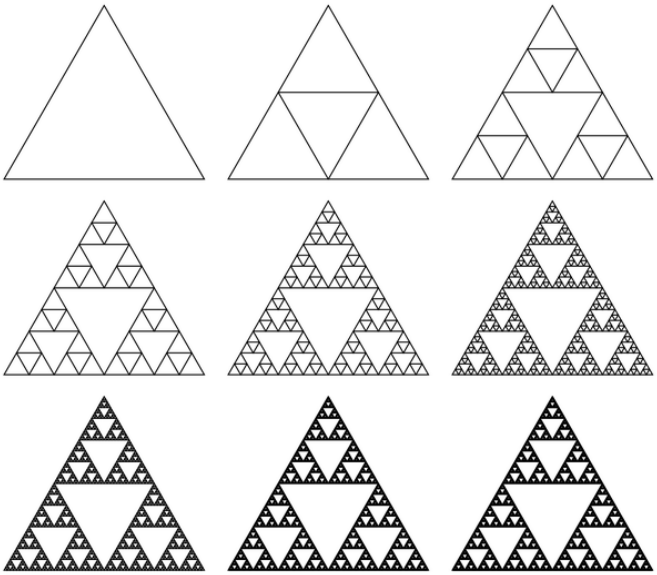

Figure 1.1: Sierpinski triangle for increasing iteratios

Scaling_Sierpinski

the scaling relation-ship in moments of order $p \in \mathbb{R}$ is derived. For $c(p) = \mathsf{E}\{|X_t|^p\}$ and $D(p) = H(p)p$ Mandelbrot et al. (1997) define a fractal process in terms of its moments, remaining graphically tractable.

$$\mathsf{E}\left(|X_t|^p\right) = c(p)t^{D(p)} \tag{1.9}$$

For normalization in p, raise the scaling law of Equation 1.9 to the power of $1/p$, giving

$$\mathsf{E}\left(|X_t|^p\right)^{1/p} = \left\{c(p)t^{D(p)}\right\}^{1/p} \tag{1.10}$$

$$\nsim \frac{1}{p}\log \mathsf{E}\left(|X_t|^p\right) = \frac{1}{p}\log c(p) + H(p)\log t. \tag{1.11}$$

If $H(p)$ is constant over p, the stochastic process is called unifractal, as one fractal dimension drives the process X_t. If $H(p)$ is varying over p, the stochastic process is called multifractal, as a spectrum of fractal dimensions drive the process X_t. What has been the Hausdorff dimension for geometric objects is the Hurst exponent for

for stochastic processes, introducing the class of self-similar processes, which are relevant for different systems in economics, biology and physics, which can be seen as stochastic.

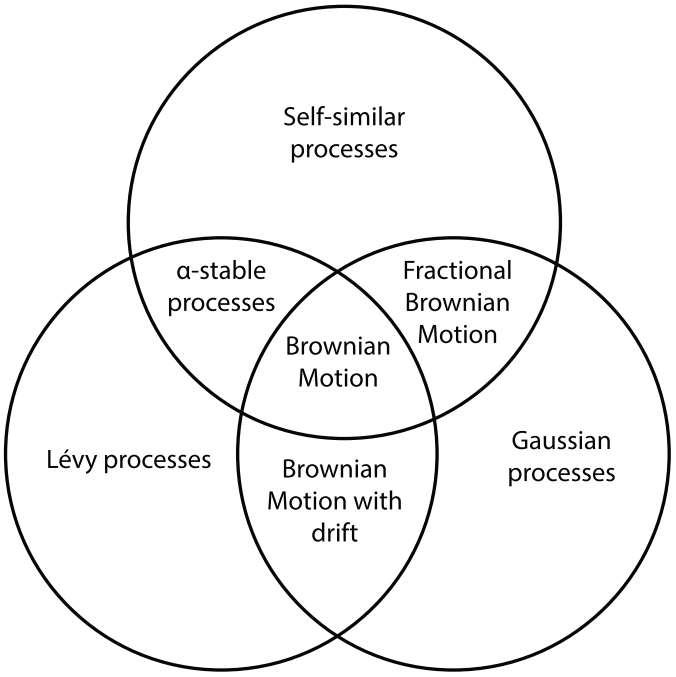

Figure 1.2: Overview over stochastic processes

1.2.1 Heavy tails through α-stabilty

One particular interesting subclass of self-similar processes are α-stable processes, which are mesofractal with Hurst exponent $H = 1/\alpha,\ p < \alpha$ and $H = 1,\ q \geq \alpha$. They are specifically interesting as stock market returns are heavy tailed over different markets, time periods and sampling frequencies. We define heavy-tailed distributions as distributions who exhibit more probability mass in the tails than under Gaussianity ($\alpha = 2$), i.e. a normalized fourth moment (Kurtosis) larger than three, including the case of infinite variance. Except for the Gaussian itself, finite variance distributions, which might be heavy tailed, change their shape un-

der aggregation. In contrast, the class of α-stable distributions is scale invariant (Mandelbrot, 1963b). Scale invariance of distribution P is defined via a continuous function g, such that for all x

$$g(\lambda)P(x) = P(\lambda x), \tag{1.12}$$

with $\lambda x \geq x_0$ and $x_0 > 0$. Equivalently, distribution P has a power-law tail, implying that for $x \geq x_0 \geq 0, c \geq 0$ and $\alpha > 0$

$$P(x) = cx^{-\alpha}. \tag{1.13}$$

In that respect, a one-dimensional random variable $X \sim \mathrm{S}(\alpha, \beta, \gamma, \delta)$ will be α-stable distributed with parameters $0 < \alpha \leq 2,\ -1 \leq \beta \leq 1,\ \gamma \geq 0$ and $\delta \in \mathbb{R}$ (Nolan, 2017; Cizek et al., 2011), if

$$X \overset{\mathcal{L}}{=} \begin{cases} \gamma Z + \delta, & \alpha \neq 1 \\ \gamma Z + (\delta + \beta \frac{2}{\pi} \gamma \log \gamma), & \alpha = 1. \end{cases} \tag{1.14}$$

$\mathrm{S}(Z \mid \alpha, \beta, 1, 0)$ represents the standard α-stable form. Only special cases of α-stable distribution are available as real-valued densities (e.g. Gaussian, Cauchy and Lévy). Scale invariance under addition implies that for the sum of α-stable variables $X_t \sim \mathrm{S}(\alpha, \beta, \gamma, \delta_t),\ t = 1, \ldots, T$

$$X_1 + X_2 + \ldots + X_T = \sum_{t=1}^{T} X_t = X \sim \mathrm{S}(\alpha, \beta, T^{\frac{1}{\alpha}}\gamma, \delta), \tag{1.15}$$

where $\delta = T\delta_t$. The summation of random variables, the convolution of densities, is crucial as lower sampling frequency returns are constructed by summing up higher frequency log-returns. Especially daily sampling frequency up to high-frequency returns exhibit heavier tails than presumed under normality are observed. McFarland et al. (1982), Boothe and Glassman (1987) and Dacorogna et al. (2001) find with increasing sampling frequency, the degree of leptocurticity increases. This is centrally reasoned by the Central Limit Theorem (CLT), which states that the

sum of the random variables X_t in time $t = 1, \ldots, T$ tends to the Gaussian as long as the first two moments of the underlying distribution are finite. Formally, let random variable X_t have expectation vector $\mu_t = \mathsf{E}(X_t)$ and covariance matrix $\sigma_t = \mathsf{E}\left[\{X_t - \mathsf{E}(X_t)\}\{X_t - \mathsf{E}(X_t)\}^\top\right]$. Then

$$\begin{aligned} \sum_{t=1}^{T} X_t &\xrightarrow{\mathcal{L}} \mathrm{N}\left(\sum_{t=1}^{T}\mu_t, \sum_{t=1}^{T}\sigma_t\right) \\ X &\xrightarrow{\mathcal{L}} \mathrm{N}(\mu, \sigma) \\ T^{-\frac{1}{2}}\sum_{t=1}^{T}(X_t - \mu_t) &\xrightarrow{\mathcal{L}} \mathrm{N}(0, \sigma)\,. \end{aligned} \tag{1.16}$$

If the distribution in horizon T is modelled as the convolution of higher frequency distributions, the multidimensional process of returns, which may not be Gaussian, but of finite variance, converges to the Gaussian. The according stochastic process with Gaussian increments is the Brownian Motion (see Figure 1.2). Accordingly, large parts of financial and also portfolio theory rest on the Gaussian assumption, see Chapter 2.

In contrast, as examined in Chapter 3, we present evidence that returns of horizons beyond the sampled frequency, are still heavy-tailed. Hence, the standard Central Limit Theorem (CLT) does not apply. Except for the Gaussian itself, finite variance distributions change their shape under aggregation. On the contrary, the class of α-stable distributions is scale invariant (Mandelbrot, 1963b). Following the results of Gnedenko and Kolmogorov (1954b), the limiting distribution of T i.i.d. α-stable random variables, $0 < \alpha \leq 2$ is

$$a_T\left(\sum_{t=1}^{T} X_t\right) - b_T \xrightarrow{\mathcal{L}} \mathrm{S}(\alpha, \beta, 1, 0), \tag{1.17}$$

where $a_T > 0$ and $b_T \in \mathbb{R}$. The special case of the Generalized Central Limit Theorem (GCLT) is the CLT of Equation 1.16 for $\alpha = 2$, $\beta = 0$, $\gamma = \frac{\sigma}{\sqrt{2}}$ and

$\delta_t = \mu_t$, given $a_T = \frac{1}{\sigma\sqrt{T}}$ and $b_T = \frac{\sqrt{T}\mu}{\sigma}$. In general, for $0 < \alpha \leq 2$,

$$T^{-\frac{1}{\alpha}} \sum_{t=1}^{T} (X_t - \delta_t) \xrightarrow{\mathcal{L}} \mathrm{S}(\alpha, 0, \gamma, 0). \tag{1.18}$$

The relevant class of α-stable processes with α-stable increments is a scale invariant Lévy process (see Figure 1.2).

1.2.2 From fractal dimension to time dependence

The Fractional Brownian Motion (FBM) is a self-similar, unifractal process with Hurst exponent $0 \leq H \leq 1$, introduced by Mandelbrot and van Ness (1968). The process $X^H(t),\ t \geq 0$ is Gaussian (see Figure 1.2), characterized by the covariance function

$$\mathsf{E}\left\{X^H(t)X^H(s)\right\} = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right). \tag{1.19}$$

For $H = 1/2$ the process collapses to the Brownian motion. For $H > 1/2$, the increments of the process are positively correlated, i.e. the process exhibits long memory. Long memory can be characterized over time lags l by

$$\sum_{l=1}^{\infty} \mathsf{E}\left\{X^H(1)(X^H(l+1) - X^H(l))\right\} = \infty. \tag{1.20}$$

When $H < 1/2$, the increments of the process are negatively correlated, i.e. the process exhibits antipersistent behavior.

1.2.3 Using self-similar stochastic processes to scale high-frequency data to lower frequencies

The standard approach in financial applications to scale a high-frequency distributions, such as daily, to a lower sampling frequency is the square-root of time rule

(Daníelsson and Zigrand, 2006). Let random variable $X_t \sim \mathrm{N}(\mu, \sigma)$ be generated by

$$X_t = \mu + \sigma Z_t, \tag{1.21}$$

where $Z_t \sim \mathrm{N}(0, 1)$ is standard Normal. By applying the CLT (see Equation 1.16), we obtain for the horizon distribution $X_T = \sum_{t=0}^{T} X_t$

$$X_T = T\mu + T^{1/2}\sigma Z_t. \tag{1.22}$$

In order to account for heavier tails for the lower sampling frequency distribution, the $T^{1/2}$-rule can be generalized under α-stable distributed random variables. Let random variable $X_t \sim \mathrm{S}(\alpha, \beta, \gamma, \delta)$ be generated by

$$X_t = \delta + \gamma Z_t, \tag{1.23}$$

where $Z_t \sim \mathrm{S}(\alpha, \beta, 0, 1)$ is standard Stable. By applying the GCLT (see Equation 1.17), we obtain for the horizon distribution $X_T = \sum_{t=0}^{T} X_t$

$$X_T = T\delta + T^{1/\alpha}\gamma Z_t. \tag{1.24}$$

In Chapter 3 we apply the $T^{1/\alpha}$-rule for portfolio returns in the context of portfolio optimization. From the point of heavy tails, we implicitly assume that the α-stable exponent is constant over the sampling frequencies. But, in Chapter 4 we find evidence for multifractality in stock market returns. On the one hand we find evidence for multifractality through changes in the Hurst exponent through the order of the absolute moments in the context of Fractional Brownian Motions (FBM). On the other hand we find varying α-stable exponents over the sampling frequency in the context of heavy tails, still differing significantly from the Gaussian assumption. By modeling the respective sampling frequencies by applying filters for intraday seasonality and long range dependence, we explain the multifractal nature of the process, giving evidence for mesofractality in the residuals. Subsequently, the

GCLT (see Equation 1.17) is utilized to scale the high-frequency residuals of the econometric time series model to the lower sampling frequency. The methodology coined Frequency Rescaling Method (FRM) is generalized to the multivariate case in Chapter 5. Serving at natural generalization of the multidimensional Brownian motion, elliptically α-stable distributions are employed and it is shown that the FRM outperforms both the $T^{1/2}$-rule and $T^{1/\alpha}$-rule.

1.3 Financial application

Investors, regulators and other financial stakeholders are rarely interested in high-frequency data, implying daily or higher. To large extents, the interests lies in weekly, monthly or yearly figures for performance, risk or stability. The relevant figures, especially estimates for risk depending on the tail of the relevant (loss) distributions, cannot be sufficiently estimated, as the data history is limited for the respective sampling frequencies or the whole data history not relevant for the profile of the object of interest. The transition of higher-frequency information to lower frequencies figures, especially related to risk, is quite often done under the Gaussian assumption implying the square-root of time rule, as argued in Subsection 1.2.3. We show that this rule of thumb is not sufficient to model the tails of stock market return distributions.

1.3.1 Portfolio theory

For portfolio optimization involving multiple financial assets, the wealth for discrete returns $X_t \in \mathbb{R}^k$ in time $t = 1, \ldots, T$, which are possible dependent random variables, is formulated as

$$W_T(f) = W_0 \prod_{t=1}^{T} \left(1 + \sum_{j=1}^{k} f_j X_{j,t} \right), \tag{1.25}$$

where $W_0 \in \mathbb{R}^+$ represents the starting wealth, $k \in \mathbb{N}^+$ is the number of assets assets with index j and $T \in \mathbb{N}^+$ are the time periods with index t. Portfolio optimization aims to chose fraction vector $f \in \mathbb{R}^k$ in order to achieve a certain criterion given potential restrictions. Accordingly, there are two main strands in the literature. On the one hand, the mean-variance approach of Markowitz (1952), Tobin (1958), Sharpe (1964) and Lintner (1965) and on the other hand the Kelly growth-optimum approach Kelly (1956), Breiman (1961) and Thorp (1971).

The Kelly betting scheme aims to maximize the expected logarithm of terminal wealth, given some underlying stochastic process for the wealth process. In Chapter 2, we show that for certain stochastic processes the Kelly growth-optimum strategy can be provided in closed-form solutions, making them easy to implement without simulations. The dealt with solutions are unconstrained solutions for the portfolio fraction, named Kelly solutions. This implies to be the strategy which cannot be asymptotically outperformed (Breiman, 1961).

The Kelly strategy is still a risky strategy which may assign putting the total wealth or even more in one investment opportunity, excluding additional risk constraints. Accordingly we restrict the Kelly Criterion with risk constraints Value at Risk (VaR) and Expected Shortfall (ES) in Chapter 3. This leads to partial Kelly solutions, which involves optimization under nonlinear constraints. Additionally, we utilize the multivariate α-stable extension of the square-root of time rule $T^{1/2}$, which generalizes to $T^{1/\alpha}$ for elliptically stable distributions. Subsequently we scale the empirical hourly NASDAQ data set to the yearly sampling frequency in order to obtain a heavy-tailed yearly distribution, providing the foundation for the constrained Kelly solution, proving the highest geometric mean. Furthermore, we show that including put options into the optimization levers the portfolio by a suitable protective put strategy, leading to an increased growth for the same level of risk.

1.3.2 Risk Management

Financial stake-holders are interested in estimating the (conditional) quantile of the wealth distribution for large confidence levels in order to report capital at risk, given a fixed probability of ungovernable events. Although it is possible to estimate risk measures for all confidence levels, given limited amount of data, we show in Chapter 4, that ten years of weekly stock market data do not suffice to estimate Value at Risk (VaR) and Expected Shortfall (ES) for confidence levels 99% and larger. From mathematical statistics it is well-known that the asymptotic distribution of the sample quantile is unbiased and Gaussian by CLT, given stationarity and finite second moments (Ruppert, 2010). As empirical time series provide a limited amount of data points, the bootstrapped sample quantile distribution are not unbiased, indicating that relevant risk measures are underestimated for low sampling frequencies such as weekly. Subsequently, we introduce the Frequency Rescaling Method (FRM), which rescales a 5-minute frequency distribution to the weekly sampling frequency by accounting for intraday seasonality, long range dependence and heavy tails in different sampling frequency domains. Accordingly, we provide a distribution for the weekly sampling frequency, which is more heavy tailed than under Gaussianity and the $T^{1/2}$-rule (Gaussian Scaling), but more platykurtic than under the $T^{1/\alpha}$-rule (Stable scaling).

We extent the Frequency Rescaling Method (FRM) to the multivariate case in Chapter 5, in which elliptically α-stable distributions are employed. Furthermore, we provide an out-of-sample test not only for high-confidence quantile estimation as in Chapter 4, but nonparametric tests for the whole weekly NASDAQ distribution, which we aim to forecast, showing the outperformance of the FRM.

In Chapter 6 we show that heavy tail behavior with the according measurements under α-stability and low/high confidence quantiles can provide a robust classification for cryptocurrencies in the domain of financial assets ranging from stocks over

exchange rates to commodities. Through means of dimensionality reduction and classification algorithms, we show that most of the variation among cryptocurrencies, stocks, exchanges rates and commodities can be explained by three factors: the tail factor, the moment factor and the memory factor. Our analysis revealed that the main difference between cryptocurrencies and classical assets, in terms of properties of the distribution of daily log-returns, is the tail behaviour, both in the left and in the right tail of the distribution, manifested by the stability index α, scale γ and high/low confidence quantiles. The moments of the distribution and the GARCH/ARCH parameters are of subliminal importance for discriminating between cryptocurrencies and classical assets. Based on the tail factor profile, we can conclude that an asset is likely to be a cryptocurrency if it has the following properties: very long tails of the log-returns distribution (in terms of the left and right quantile and the conditional tail expectation), high variance, high value of the α-stable scale parameter and a value of the α-stable tail index closer to 1, reflecting the Cauchy distribution.

From the point of view of the risk analysts and regulators, the non-linear classification applied on the factors extracted, provide proficient results in order to discriminate between cryptocurrencies and the other asset classes. Through means of an expanding window approach, we are able to depict the evolutionary dynamics of the cryptocurrency universe and show how the clusters are formed by projecting the multidimensional dataset on the main factors converging over time. By interpreting the asset universe as a complex ecosystem, we are able to depict that cryptocurrencies exhibit a synchronic evolution, implying that individual cryptocurrencies develop similar characteristics over time, on the one hand. On the other hand we observe a divergent evolution, cryptocurrencies as different species, compared to the classical asset classes.

Chapter 2

The Kelly Criterion: A probabilistic perspective

2.1 Introduction

For portfolio optimization given multiple financial assets, the wealth for discrete returns $X_i \in \mathbb{R}^k$, which are possible dependent random variables, is formulated as

$$W_n(f) = W_0 \prod_{i=1}^{n} \left(1 + \sum_{j=1}^{k} f_j X_{j,i}\right). \tag{2.1}$$

$W_0 \in \mathbb{R}^+$ represents the starting wealth, $k \in \mathbb{N}^+$ is the number of assets assets with index j and $n \in \mathbb{N}^+$ are the time periods with index i.

Portfolio optimization implies to chose fraction vector $f \in \mathbb{R}^k$ in order to achieve a certain growth criterion given potential risk constraints. Accordingly, there are two main strands in the literature. On the one hand, the mean-variance approach of Markowitz (1952), Tobin (1958), Sharpe (1964) and Lintner (1965) and on the other hand the Kelly growth-optimum approach of Kelly (1956), Breiman (1961) and Thorp (1971).

The Kelly betting scheme aims to maximize the expected logarithm of terminal

wealth, given some underlying stochastic process for the wealth process. For certain stochastic processes the Kelly growth-optimum strategy can be provided in closed-form solutions, making them easy to implement without simulations. The dealt with solutions are unconstrained portfolio fraction named Kelly solutions. This implies to be the strategy which cannot be asymptotically outperformed (Breiman, 1961). The Kelly strategy might still be a risky strategy assigning to put the total wealth or even more in one investment opportunity, excluding additional risk constraints. Risk constraints, such as Value at Risk (VaR) oder Expected Shortfall (ES) as part of the general optimization problem will be dealt with in a separate chapter, leading to partial Kelly solutions.

Shortly reviewing the central results of Kelly (1956), the situation under binary events is extended by including uneven odds and a minimum bet in the expected growth rate (Thorp, 1984). Whereas Bicksler and Thorp (1973) extend the approach of Kelly to uniform returns, Merton (1992) derives, under the assumption of the geometric Brownian Motion for price changes, a continuous life time portfolio strategy. Due to the assumption that prices are log-normally distributed, the first two moments describe the distribution sufficiently. The estimation of expectation and variance leads, as the true distribution parameters are unknown, inevitable to errors, which are examined by Chopra and Ziemba (1993) in a relative manner to each other, indicating that the mean represents the core driver of portfolio performance, followed by variances and covariances. Thorp (2006) shows that the results hold for the investment fraction when the outcome follows no specific distribution, but mean and variance are finite. The wealth process is approximated by a second-order Taylor approximation in order to obtain a closed-form solution. As especially financial returns are heavier tailed than presumed under Gaussianity, Osorio (2008) generalizes the result of Merton (1992) for the Student-T distribution. Lv and Meister (2009) apply the Kelly criterion under multiple Ornstein-Uhlenbeck processes, whereas MacLean et al. (2005) review the

growth-optimum solution in a hierarchical Bayesian framework under Gaussianity.

2.2 Bernoulli trials (Kelly, 1956)

Kelly (1956) introduced the log-utility function in the context information theory and implicitly gambling. Given Bernoulli trials (Binary Channel) he shows that the use of the logarithmic utility function maximizes the long run growth rate. In the short-term he points out that the utility function is myopic by period-by-period maximization. The fraction only rests upon the current value of initial capital.
Suppose a favorable game with winning probability $\frac{1}{2} < p \leq 1$ and outcome 1 and losing probability $q = 1 - p$ with outcome -1 (even odds), starting with initial wealth W_0. The wealth after $n \in \mathbb{N}^+$ trials, out of which $m \leq n$ are won, given betting fraction $0 < f \leq 1$ of the initial capital (in %), is given by

$$W_n = W_0(1+f)^m(1-f)^{n-m}. \tag{2.2}$$

Maximizing the expectation of the wealth process W_n implies for $\mathrm{P}(X = 1) = p = 1$, that the investor bets all of his wealth, $f = 1$, leading to

$$W_n = W_0 2^n. \tag{2.3}$$

Under uncertainty, $p < 1$, maximizing the expectation of wealth still implies $f = 1$,

$$\mathsf{E}(W_n) = W_0 + \sum_{i=1}^{n} (p-q)\,\mathsf{E}\left(fW_{n-1}\right), \tag{2.4}$$

which leads to ruin asymptotically

$$\mathrm{P}\left(\{W_n \leq 0\}\right) = \mathrm{P}\left\{\lim_{n\to\infty}\left(1-p^n\right)\right\} \to 1. \tag{2.5}$$

Minimizing the probability of ruin alternatively is achieved for $f = 0$, as

$$\mathrm{P}\left(\{W_n \leq 0\}\right) = 0. \tag{2.6}$$

Accordingly, the minimum ruin strategy also leads to the minimization of the expected profits as no investment takes place.
Kelly (1956) aims to maximize he exponential rate of asset growth per trial, equalling the logarithm of the geometric mean.

$$\begin{aligned} G_n(f) &= \log\left(\frac{W_n}{W_0}\right)^{\frac{1}{n}} = \log\left\{(1+f)^{\frac{m}{n}}(1-f)^{\frac{n-m}{n}}\right\} \\ &= \left(\frac{m}{n}\right)\log(1+f) + \left(\frac{n-m}{n}\right)\log(1-f). \end{aligned} \tag{2.7}$$

Consequently, the expected growth rate coefficient is given by

$$\mathsf{E}\left\{G_n(f)\right\} = g(f) = p\log(1+f) + q\log(1-f), \tag{2.8}$$

where p is the winning probability and $q = 1-p$ the losing probability. Maximizing $g(f)$ with respect to f leads to

$$\begin{aligned} g'(f) &= \left(\frac{p}{1+f}\right) - \left(\frac{q}{1-f}\right) = \left\{\frac{p-q-f}{(1+f)(1-f)}\right\} = 0 \\ &\Leftrightarrow f = f^* = p - q > 0, \quad p > q. \end{aligned} \tag{2.9}$$

The second derivative according to f shows that $f = f^*$ is the unique maximum of the function $g(f^*) = p \cdot \log(p) + q \cdot \log(q) + \log(2) > 0$.

$$g''(f) = -\left\{\frac{p}{(1+f)^2}\right\} - \left\{\frac{q}{(1-f)^2}\right\} < 0 \tag{2.10}$$

Theorem 2.2.1 (Kelly)**.** *The optimal fraction, under Bernoulli trials, which should be invested per trial, is $f^* = p - q$, the edge. This fixed fraction strategy maximizes the expected value of the logarithm of capital at each trial (Kelly, 1956).*

As Thorp (1971) points out later, maximizing the expected logarithm of wealth $\mathsf{E}\left\{\log(W_t)\right\}$ is equivalent to maximizing the exponential rate of growth per time period $g(f)$.

2.2.1 Even odds - Kelly (1956)

Figure 2.1 plots the surface of the expected growth rate as a function of the fraction and the winning probability. The green parts of the surface are fraction-probability combinations through which $g(f, p)$ remains positive. The four data tips represent the optimal sets of portfolios for $p = [0.5, 0.6, 0.8, 0.95]$. The optimally invested fraction is a linear function on the winning probability.

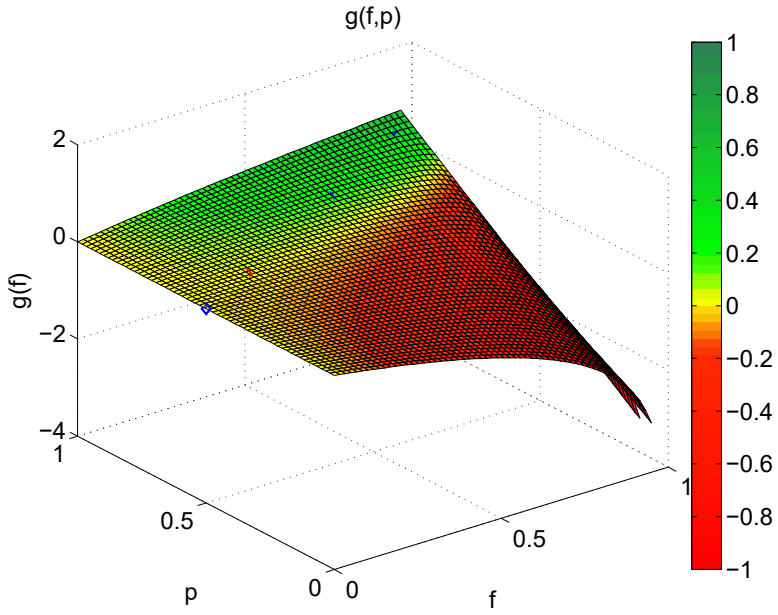

Figure 2.1: Logarithm of the geometric growth rate $g(f, p)$ depending on fraction f and winning probability p, data tips visualize the Kelly fraction $f^* = [0, 0.2, 0.6, 0.9]$ for $p = [0.5, 0.6, 0.8, 0.95]$

Kelly_Bernoulli

Plotting the expected growth rate as a function of the fraction f with known winning probability, the surface $g(f, p)$ reduces to the function $g(f)$ for $p = 0.6$, which shall be maximized (Figure 2.2). Optimization algorithms aim to minimize convex or non-convex functions, so the location of the maximum of $g(f)$ is the location of the minimum of -$g(f)$. For illustration purposes the original $g(f)$ was

plotted and the maximum was marked with a data tip.

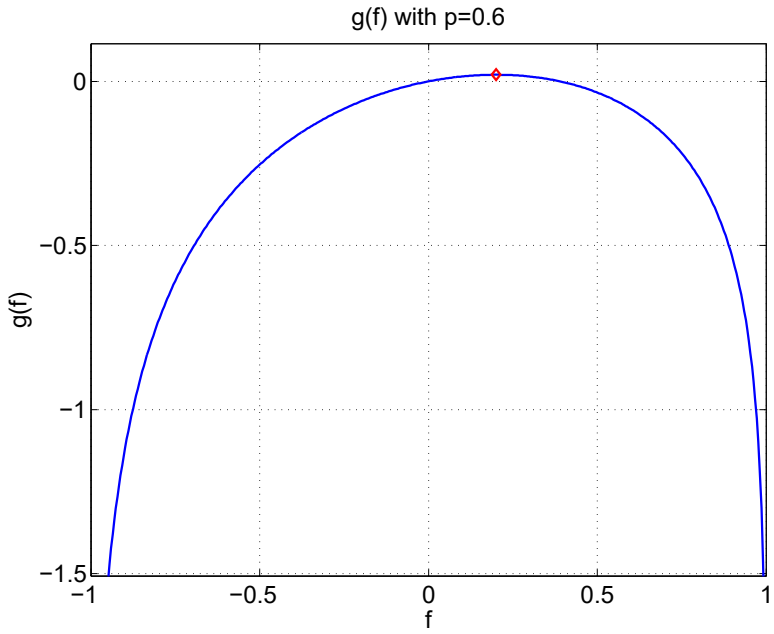

Figure 2.2: Logarithm of the geometric growth rate $g(f)$ depending on fraction f with fixed winning probability at $p = 0.6$, data tip for the optimal fraction $f^* = 0.2$

Kelly_Bernoulli

Moreover, Figure 2.2 indicates that the expected growth rate cannot be negative if $f \leq f^*$ for $f \geq 0$. But if the winning probability is significantly overestimated, $g(f)$ becomes negative. This is also a starting point for risk-averse partial Kelly solutions.

2.2.2 Uneven odds - Thorp (1984)

Assume that the odds are not even anymore, so $o \in \mathbb{R}^+$.[1] Thereafter, a game is favorable if $po - q > 0$, leading to a variation of the logarithm of the geometric

[1] If the odds for winning are 1:1, the payoffs for both events, winning and losing, are even. If the odds are for example 2:1, the probability of losing is two times the probability of winning, so the payoff for winning is two times the payoff for losing.

growth rate

$$go(f, o) = p \log(1 + of) + q \log(1 - f), \tag{2.11}$$

which is maximized, using ordinary calculus, by

$$f^* = \frac{op - q}{o} = \frac{\text{edge}}{\text{odds}}. \tag{2.12}$$

From the analytical solution, $\partial f^*/\partial o > 0$ and $\partial^2 f^*/\partial o^2 < 0$. In order to examine the effect of changing odds on the fraction, the surface of the function $go(f, o)$ with $p = 0.6$ is plotted in Figure 2.3. The five data tips represent the optimal sets of portfolios for $o = [1, 2, 3, 4, 5]$. The invested fraction is not linear in the odds.

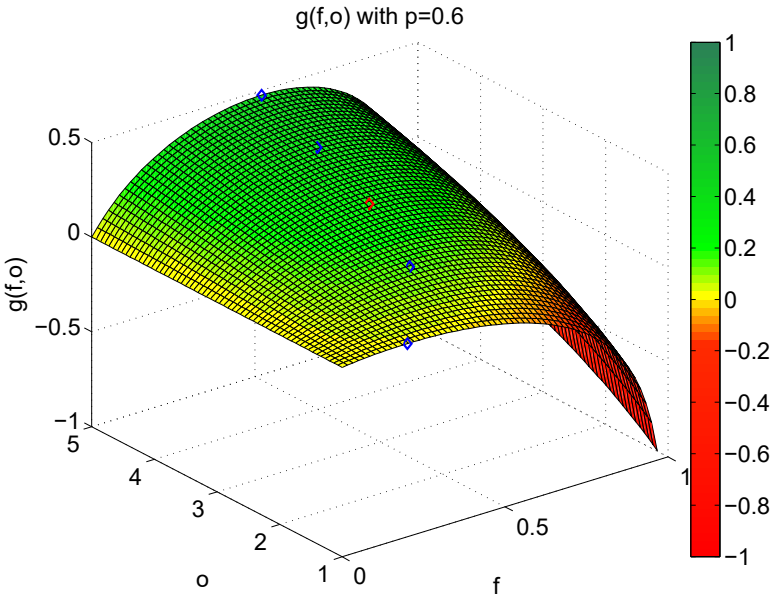

Figure 2.3: Logarithm of the geometric growth rate $g(f, o)$ depending on fraction f and odds o with fixed winning probability $p = 0.6$, data tips for the Kelly fraction with varying odds $o = [1, 2, 3, 4, 5]$

Thus, fixing $o = 3$ leads to the maximization of $g(f)$ given $o = 3$ and $p = 0.6$ (Figure 2.4). The observation from Figure 2.2, that solely overaggressive betting leads to a negative expected growth rate, holds.

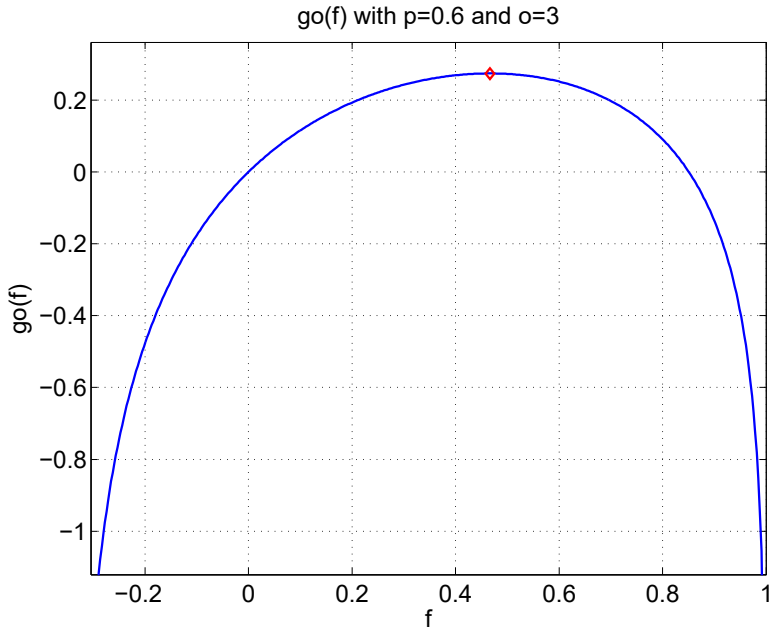

Figure 2.4: Logarithm of the geometric growth rate $g(f)$ depending on fraction f with odds $o = 3$ and winning probability $p = 0.6$ data tip for the Kelly fraction

2.2.3 Even odds and a minimum bet - Thorp (2006)

There are many games such as Blackjack or Poker, in which it would be seen curious, or where it is impossible, if one would play solely favorable situations. Hence, there is often a minimum bet $a \in [0, 1]$ involved. Let f be the bet on the favorable situation, where an edge is given, and af be the bet on unfavorable situations. Using $\mathrm{P}(x)$ as notation for the probability for the favorable game situation, Thorp (2006) modifies the Kelly growth rate in the following way:

$$\begin{aligned} ga(f, a, \mathrm{P}(x)) = \mathrm{P}(x) \left\{ p \log(1 + f) + q \log(1 - f) \right\} \\ + \left\{ 1 - \mathrm{P}(x) \right\} \left\{ q \log(1 + af) + p \log(1 - af) \right\}. \end{aligned} \tag{2.13}$$

To visualize the function which is going to be be maximized, we restrict on $\mathrm{P}(x)$ and the winning probability, which we assume to be given in a certain game. Changing the expected growth rate from Thorp (2006) by

i. approximating $\mathrm{P}(x)$ with $(1/n)$, where $n \in \mathbb{N}^+$ is the number of players,

ii. assuming two players, implies $\mathrm{P}(x) = 0.5$ and

iii. an edge of $p - q = 0.2$

leads to the maximization of the expected growth rate

$$\begin{aligned} ga(f,a) = 0.5\left\{0.6\log(1+f) + 0.4\log(1-f)\right\} \\ +0.5\left\{0.4\log(1+af) + 0.6\log(1-af)\right\}. \end{aligned} \tag{2.14}$$

In the first place, the surface for the expected growth rate $g(f,a)$ was plotted as a function of the fraction f and the minimum bet a as a percentage of the fraction (Figure 2.5). The four data tips represent the optimal fractions for $a = [0, 0.05, 0.2, 0.5]$. The Kelly gambler reduces the optimal fraction, starting with no minimum bet, from 0.2 to zero as a tends to one (Thorp, 2006). From the practical point of view, under approximation of $\mathrm{P}(x)$, the fractions with varying a should be calculated beforehand, due to the fact that the maximization $ga(f,a)$ has to be done numerically. For $a = 0.05$, the surface $ga(f,a)$ reduces to the function $ga(f)$ with $p = 0.6$, which was plotted with a data tip at the maximum of the function (Figure 2.6).

Moreover, it is important to indicate that the Kelly fraction further reduces when $\mathrm{P}(x)$ decreases, respectively when the number of players increases (see Table 2.1).

Players	a	0	0.2	0.4	0.6	0.8	1
2	f^*	0.2	0.155	0.104	0.059	0.024	≈ 0
3	f^*	0.2	0.112	0.03	≈ 0	≈ 0	≈ 0
4	f^*	0.2	0.072	≈ 0	≈ 0	≈ 0	≈ 0

Table 2.1: Optimally invested fraction depending on the number of players and the minimum bet a

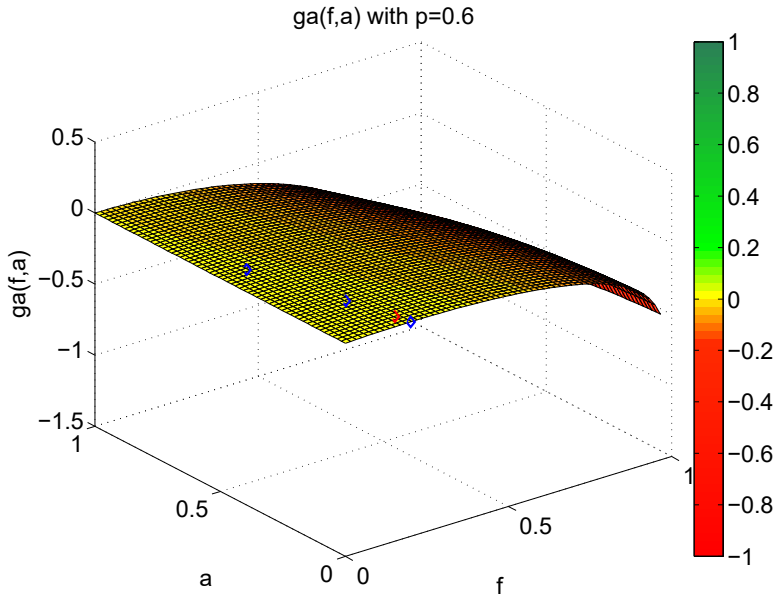

Figure 2.5: Logarithm of the geometric growth rate $g(f, a)$ depending on fraction f and minimum bet a with fixed winning probability $p = 0.6$, data tips for the Kelly fraction with varying minimum bets $a = [0, 0.05, 0.2, 0.5]$

2.3 Uniform returns (Bicksler and Thorp, 1973)

Presume a market of one risky asset, where the return is uniformly distributed on lower bound a and upper bound b, so $x \sim U(a, b),\ a < b$. Additionally, the investor can buy a risk free asset with constant return r. So, the wealth W_n can be written as

$$W_n = W_0 \left\{1 + r + f(x - r)\right\}. \tag{2.15}$$

The exponential growth rate for this opportunity is

$$G(f) = \log\left(\frac{W_n}{W_0}\right) = \log\left\{1 + r + f(x - r)\right\} \tag{2.16}$$

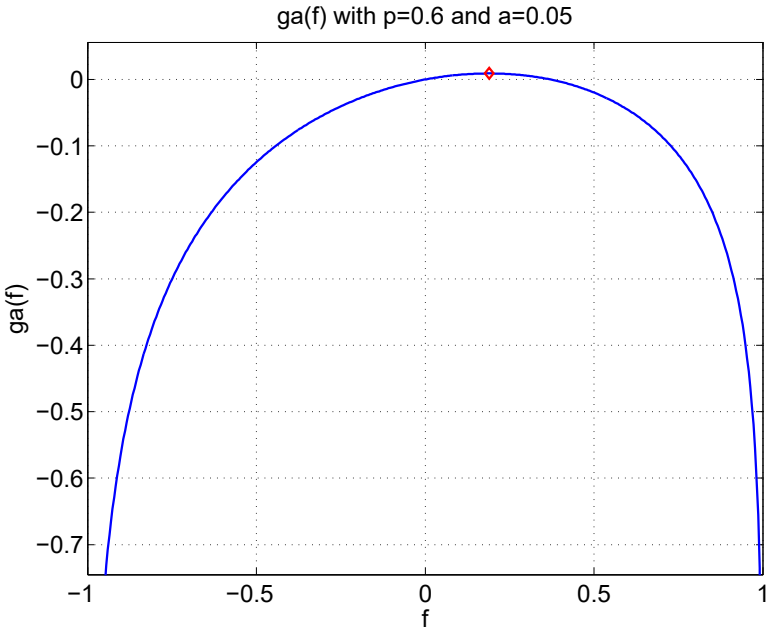

Figure 2.6: Logarithm of the geometric growth rate depending on fraction f with fixed minimum bet $a = 0.05$ and winning probability $p = 0.6$, data tip for the Kelly fraction

and the maximization problem can be formulated as

$$\begin{aligned} f^* &= \arg\max_f \mathsf{E}\left\{G(f)\right\} = \arg\max_f g(f) \\ &= \arg\max_f \mathsf{E}\left\{\log\left[1 + r + f(x - r)\right\}\right]. \end{aligned} \tag{2.17}$$

The first order condition is given and rewritten by

$$\begin{aligned} &\int_a^b \left\{\frac{x - r}{1 + r + f(x - r)}\right\} \times \left\{\frac{1}{b - a}\right\} dx = 0 \\ \Leftrightarrow\ & f(b - a) = (1 + r)\log\left\{\frac{1 + r + f(b - r)}{1 + r + f(a - r)}\right\} \\ \Leftrightarrow\ & \left\{\frac{1 + r + f(b - r)}{1 + r + f(a - r)}\right\}^{1/f} = \exp\left\{\frac{b - a}{1 + r}\right\}. \end{aligned} \tag{2.18}$$

As seen, there is no suitable closed form solution, so the optimal fraction f has to be calculated using numerical procedures such as Newton Raphson or Bisection method.

For exemplary analysis the bounds of the uniform r.v. x are set to $[a, b] = [-0.5, 0.5]$ and the risk free rate to $r = 0.01$. Following the Kelly strategy for those parameters, would imply to invest all capital into the risky asset as the distribution is symmetric aroud zero. Letting the upper bound b of the r.v. increase from 0.01 to 1, the optimal fraction increases non-linearly (Figure 2.7). Letting the lower bound a of the r.v. increase from 0.01 to 1, the optimal fraction decreases non-linearly (Figure 2.8).

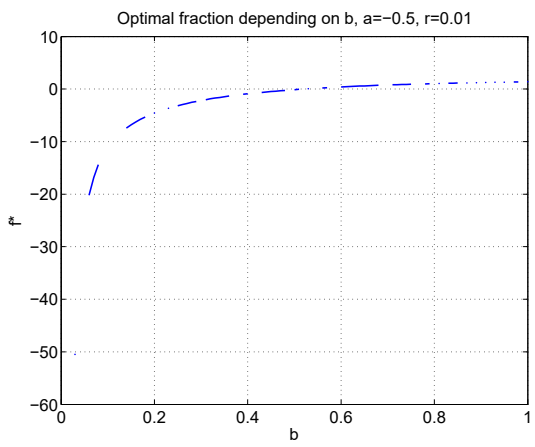

Figure 2.7: Optimally invested fractions under uniform with changing bounds $[-0.5, b]$

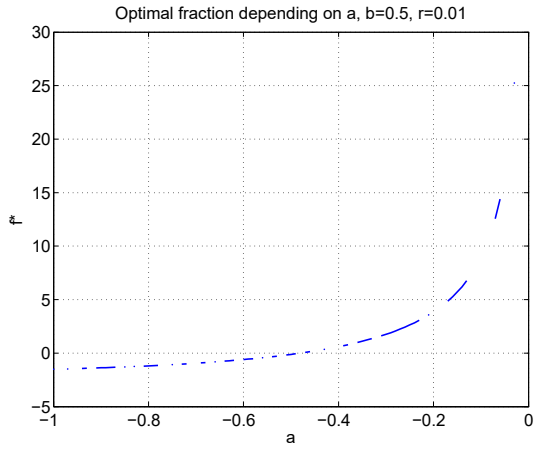

Figure 2.8: Optimally invested fractions under uniform with changing bounds $[a, 0.5]$

2.4 Log-normal prices (Merton, 1969/1992)

The goal of this section is to derive a closed-form solution for the optimal fraction under lognormal prices P_j, $j = 1, \ldots, k$ for $k \in \mathbb{N}^+$ assets, Gaussian log-returns $X_j \sim \mathrm{N}(\mu_j, \sigma_j^2)$. The classic continuous-time solution of the inter-temporal investment-consumption problem is mainly due to Merton (1969) and has been extended by many researchers, e.g. Browne (1997) and Browne (2000). The focus

primary lies on Chapter four of his book "Continuous time finance" (Merton, 1992). The crucial assumption, in order to derive the following results, is that the logarithm of the price ratio, $\log\left(\frac{P_{j,t+1}}{P_{j,t}}\right) = \log(X_{j,t})$, follows a Geometric Brownian Motion (GBM), also called Itô-process. Accordingly, the price of the risky asset j satisfies the stochastic differential equation

$$dP_{j,t} = \mu_{j,t} P_{j,t} dt + \sigma_{j,t} P_{j,t} dZ_{j,t}, \tag{2.19}$$

where $Z_{j,t}$ are standard Brownian Motions, which can be dependent. Additionally, a risk free asset with price R and risk free return $0 \leq r < \mu_j$ is assumed, evolving according to

$$dR_t = r R_t dt. \tag{2.20}$$

Consistent with the Black-Scholes-Merton approach, the parameters μ_j, σ_j and r are supposed to be constants - fixed over time - to attain a time-constant solution. The continuous wealth process, depending on the consumption C in period t, can be described as

$$dW_t = \left\{ \sum_{j=1}^{k} f_{j,t} \mu_j W_t \right\} dt + \sum_{j=1}^{k} f_{j,t} \sigma_j W_t dZ_{j,t}. \tag{2.21}$$

For the univariate case, one risky and one risk-free asset, the wealth dynamics can be rewritten in the following form:

$$dW_t = \left\{ (f\mu + (1-f)r) W_t - C_t \right\} dt + f\sigma W_t dZ_t. \tag{2.22}$$

The lifetime objective function (Merton, 1992) is given by

$$I\left\{W_t, t\right\} = \max_{C,f} \mathsf{E}\left\{ \int_0^T e^{-\rho t} U(C_t) dt + B(W_T, T) \right\}, \tag{2.23}$$

with impatience factor ρ and the Bequest valuation function at time T, concave in wealth at T. Using a Taylor approximation at t and taking expectations

$$0 = \max_{C,f} \left[e^{-\rho t} U(C_t) + \frac{\partial I(W_t,t)}{\partial t} + \frac{\partial I(W_t,t)}{\partial W} \left\{ f_t((\mu - r) + r)W_t - C_t \right\} + \frac{1}{2} \frac{\partial^2 I(W_t,t)}{\partial W^2} f_t^2 \sigma^2 W_t^2 \right] \equiv \phi \tag{2.24}$$

with first order conditions

$$\begin{aligned} \phi_c &= e^{-\rho t} U'(C^*) - \frac{\partial I(W_t,t)}{\partial W} = 0, \\ \phi_w &= (\mu - r) W \frac{\partial I(W_t,t)}{\partial W} + \frac{\partial^2 I(W_t,t)}{\partial W^2} f^* W^2 \sigma^2 = 0. \end{aligned} \tag{2.25}$$

The solution to ϕ, the live time objective function is not trivial; hence, it needs to be simplified. Assume that

$$J(W_t,t) = e^{-\rho t} I(W_t,t). \tag{2.26}$$

Letting $T \to \infty$ the Bequest function at T, $B(W_t,T)$, falls out. The new objective function can be written as

$$J\{W_t\} = \max_{C,f} \mathsf{E} \left\{ \int_0^\infty e^{-\rho t} U(C_t) dv \right\}, v \in [0,\infty]. \tag{2.27}$$

Thus, the Partial Differential Equation (PDE) simplifies to the Ordinary Differential Equation (ODE)

$$0 = \max_{C,f} \left[U(C_t) - \rho J(W) + \frac{\partial J(W_t,t)}{\partial W} \left\{ f_t((\mu - r) + r)W_t - C_t \right\} + \frac{1}{2} \frac{\partial^2 J(W_t,t)}{\partial W^2} f_t^2 \sigma^2 W_t^2 \right], \tag{2.28}$$

which is no longer a function of time, due to the fact that dt fell out. Poon (2010) describes this step as a "key development in solving the life time consumption decision".

Remembering Thorp (1971), the main aim is to present optimal portfolio strategies under the log-utility function in a normative way. For the case of CRRA (Constant

Relative Risk Aversion). The isoelastic marginal utility is given by

$$U(C) = \frac{1}{\gamma} C^{\gamma}, \tag{2.29}$$

with relative risk aversion (RRA)

$$RRA = -\frac{U''(C)}{U'(C)} C = \frac{-(\gamma - 1) C^{\gamma - 2}}{C^{\gamma - 1}} C = 1 - \gamma, \tag{2.30}$$

which is a constant, therefore, constant RRA. If $U(C) = \log(C)$, then $\gamma = 0$ and $RRA = 1$. For the isoelastic case, substituting the RRA into the the FOC ϕ_c gives

$$\begin{aligned} e^{-\rho t}(C^*)^{\gamma - 1} = I'(W) \Leftrightarrow C^* = \left\{ e^{-\rho t} I'(W) \right\}^{\frac{1}{\gamma - 1}}, \\ f^* = -\left(\frac{\mu - r}{\sigma^2} \right) W \frac{J'(W)}{J''(W)}. \end{aligned} \tag{2.31}$$

For the $T \to \infty$, the optimal decision rules can be rewritten as

$$C^* = J'(W)^{\frac{1}{\gamma - 1}}, \tag{2.32}$$

$$f^* = -\left(\frac{\mu - r}{\sigma^2} \right) W \frac{J'(W)}{J''(W)}. \tag{2.33}$$

Following the solution of J(W) with $\frac{v^{\gamma - 1}}{\gamma} W^{\gamma}$ the following theorem can be stated.

Theorem 2.4.1. *Univariate Solution*

Assuming that the infinitesimal price changes follow a GBM under an isoelastic marginal utility $U(C) = \frac{1}{\gamma} C^{\gamma}$ with CRRA, Merton (1969) shows that the optimal consumption and investment rules in the infinite time case are given by

$$C^*_{\infty,t} = \left\{ \frac{\rho}{1 - \gamma} - \gamma \left(\frac{(\mu - r)^2}{2\sigma^2 (1 - \gamma^2)} + \frac{r}{1 - \gamma} \right) \right\} \tag{2.34}$$

$$f^*_{\infty} = \left(\frac{\mu - r}{\sigma^2} \right) (1 - \gamma). \tag{2.35}$$

The optimal consumption and investment strategy is, consistent with e.g. Hakansson (1970) or Hakansson (1971), independent of wealth and consumption. Owing to the fact, that μ, σ and r are supposed to be constants, the optimal fraction f^*_{∞} solely depends on the risk aversion parameter γ.

Corollary 2.4.1. *Log-utility*

Assuming the logarithmic utility, so $\gamma = 0$*, theorem* 1 *simplifies to*

$$C^*_{\infty,t} = (\rho + r)W_t \tag{2.36}$$

$$f^*_\infty = \left(\frac{\mu - r}{\sigma^2}\right). \tag{2.37}$$

Consumption becomes a linear function of wealth and the invested fraction exhibits a close relationship to the Sharpe-Ratio (Poon, 2010).

Theorem 2.4.2. *Multivariate Solution*

Consequently, Merton (1992) extends the univariate solution to the multivariate analogue, given by

$$f^*_\infty = \Sigma^{-1}(\mu - r1), \tag{2.38}$$

with the fixed fraction vector $f^*_\infty = \begin{bmatrix} f^*_{j,\infty} \\ \vdots \\ f^*_{k,\infty} \end{bmatrix}$, the covariance matrix

$\Sigma = \begin{bmatrix} \sigma_j^2 & \dots & \sigma_{j,k} \\ \vdots & \ddots & \vdots \\ \sigma_{j,k} & \dots & \sigma_k^2 \end{bmatrix}$, containing the variances of assets $j : k$ in the diagonal and the covariances around, the mean vector $\mu = \begin{bmatrix} \mu_j \\ \vdots \\ \mu_k \end{bmatrix}$ and the risk free rate r multiplied by an appropriate vector of ones $1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$.

It is important to consider is that the results under the GBM-assumption, starting from Breiman (1961), up to Algeot and Cover (1988), hold also for Merton's continuous time solution, although Merton never relates to the names Kelly or

Breiman. On the one hand, Merton rests his solution on the maximization of the expected logarithm of terminal capital and therefore indirectly follows Kelly (1956), but on the other hand he widens the field of utility functions and provides more general solutions.

2.5 A continuous approximation (Thorp, 2006)

In order to apply the Kelly Criterion for security markets, continuous distributions are relevant. As reasoned before, the goal is to maximize $g(f) = \mathsf{E}\{\log(1+fx)\} = \int \log(1+fx)d\,\mathrm{P}(x)$ with $\mathrm{P}(x)$ as probability measure for the outcomes and f as invested fraction of capital, which we aim to optimize. Constraints are $1+fx>0$, so log is defined and $\sum f_j = 1$. If we let outcomes x be a symmetric r.v. around $\mathsf{E}(x) = \mu$ with $Var(x) = \sigma^2$, the wealth W can be described as

$$W(f) = W_0\{1+(1-f)r+fx\} = W_0\{1+r+f(x-r)\}, \tag{2.39}$$

with r as return of the risk free asset. Consequently,

$$g(f) = \mathsf{E}\{G(f)\} = \mathsf{E}\log\{W(f)/W_0\} = \mathsf{E}\log\{1+r+f(x-r)\}. \tag{2.40}$$

For a subdivided time interval with T independent steps

$$\frac{W_T(f)}{W_0} = \prod_{t=1}^{T}\{1+(1-f)r+fx_t\}. \tag{2.41}$$

Taking expectation and natural logarithm on both sides leads to $g(f)$. This result derives from a second order Taylor-approximation:

$$g(f) = r + f(\mu - r) - \sigma^2 f^2/2 + \mathcal{O}(n^{-1/2}). \tag{2.42}$$

For $t \longrightarrow \infty$, $\mathcal{O}(n^{-1/2}) \longrightarrow 0$, leading to

$$g_\infty(f) = r + f(\mu - r) - \sigma^2 f^2/2. \tag{2.43}$$

Differentiating $g(f)$ according to f leads to

$$\frac{\partial g_\infty(f)}{\partial f} = \mu - r - \sigma^2 f = 0, \tag{2.44}$$

giving

$$f^* = \frac{\mu - r}{\sigma^2}. \tag{2.45}$$

The result holds for any bounded r.v. with the first two moments μ and σ^2. Note that this is the same fraction as derived in Merton (1992), assuming the log-utility function in the normative sense. Thorp (2006) observes that as $t \longrightarrow \infty$, the wealth tends to a log-normal diffusion process with an underlying security with drift μ and variance rate σ^2. So, $g_\infty(f)$ is the instantaneous growth rate of depending on fraction f:

$$g_\infty(f) = r + f(\mu - r) - \sigma^2 f^2/2. \tag{2.46}$$

Betting the optimal fraction f^* leads to growth rate

$$g_\infty(f^*) = \frac{(\mu - r)^2}{2\sigma^2} + r. \tag{2.47}$$

For the given approximation, $g_\infty(f)$ is parabolic around f^* with range $0 \leq f^* \leq 2f^*$. From the log-normality of $W(f)/W_0$ it follows that

$$\begin{aligned} \log\left(\frac{W(f)}{W_0}\right) &\sim \mathrm{N}(\tilde{\mu}t, \tilde{\sigma}^2 t), \\ &\sim \mathrm{N}(g_\infty(f)t, \sigma^2 f^2 \cdot t). \end{aligned} \tag{2.48}$$

Furthermore, Thorp solves for the expected portfolio growth with scaling k as $t_k g_\infty = k t_k^{\frac{1}{2}} Std(G_\infty(f))$ to get

$$t_k = \frac{k^2 \sigma^2 f^2}{g_\infty^2}. \tag{2.49}$$

Thorp (2006) points out that the moments of x and also the risk free rate r are changing over time, leading to a changing optimal fraction f^*. Without further detail he proposes to re-estimate the optimal fraction periodically.

The multivariate case is derived,in accordance with Merton (1992), analogously:

$$f^* = \Sigma^{-1}(\mu - r1) \tag{2.50}$$

$$g_\infty(f^*) = r + \frac{f^{*\top}\Sigma f^*}{2}. \tag{2.51}$$

Exemplary, assume that $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ with $\mu_1 = 0.04$, representing an annual asset growth rate of 4% and $\sigma_1 = 0.15$, giving the annual standard deviation. A risk free rate of $r = 0.01$ is presumed. Accordingly the growth-optimal fixed fraction strategy would be to invest $f^* = 1.33$ in the risky asset, implying a short position of -33% in the risk free asset.

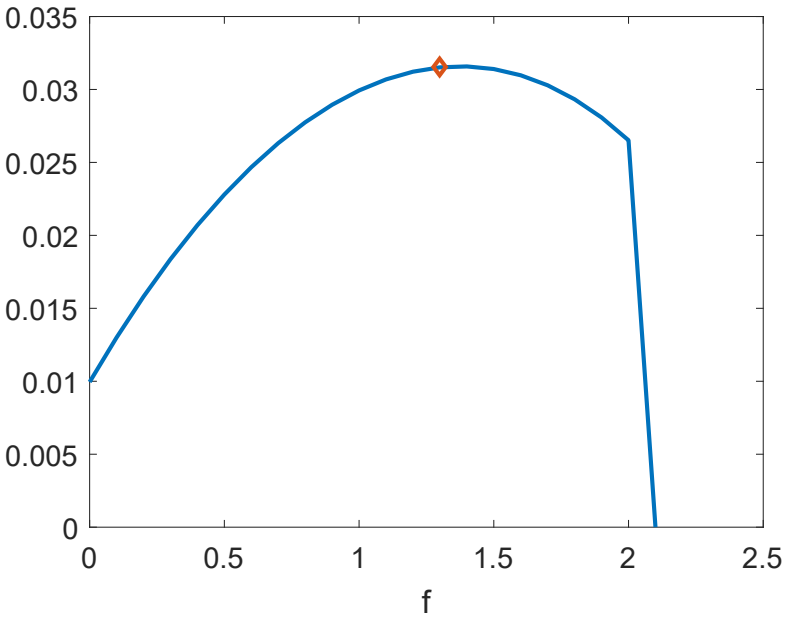

Figure 2.9: Logarithm of the geometric growth rate $g(f)$ under Gaussianity, $X \sim \mathrm{N}(0.04, 0.15^2)$, with Kelly fraction at $f^* = 1.33$

Kelly_Gaussian

Adding one uncorrelated asset ($\rho = 0$) with the same parameters, except having a higher expectation of $\mu_2 = 0.08$, so 8% p.a., implies $X \sim \mathrm{N}(\mu, \Sigma)$ with mean vector $\mu = [0.04, 0.08]^\top$ and covariance matrix $\Sigma = \begin{bmatrix} 0.15^2 & 0 \\ 0 & 0.15^2 \end{bmatrix}$.

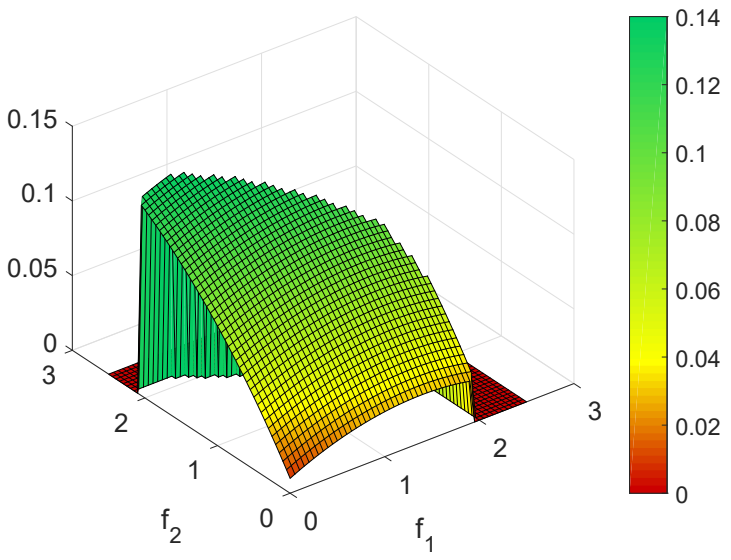

Figure 2.10: Logarithm of the geometric growth rate $g(f_1, f_2)$ under Gaussianity, $X \sim \mathrm{N}(\mu, \Sigma)$

Kelly_Gaussian

2.6 Student-T returns (Osorio, 2008)

Osorio (2008) argues that especially stock prices are not log-normally distributed as argued in Merton (1992). Excess kurtosis and skewness cannot be sufficiently captured. Having a daily return x, coming from the probability measure $\mathrm{P}(x)$, the Kelly bet implies betting the fraction f^*, which maximizes $\mathsf{E}\{\log(1 + fx)\}$.

For the continuous return distribution $\mathrm{P}(x)$ in the $(-\infty, \infty)$ domain, the aim is to maximize

$$\mathrm{U}(f) = \int_{-\infty}^{\infty} \mathrm{P}(x) \log(1 + fx) dx. \tag{2.52}$$

In order to avoid that the log function becomes zero or negative, the integral requires a lower bound $x_1 > -1/h$. For distributions decaying sufficiently fast to zero, the truncation does not imply to cut off significant probability mass. But for heavy-tailed, slower decaying distributions this approximation may not

be sufficient. Osorio (2008) alternates the problem formulation for unbounded probability in two steps:

i. Specify a small number $\delta << 1$, representing the tail area, which should be neglected by the log-utility. x_1 and x_2 are the according thresholds in the sense of

$$\int_{-\infty}^{x_1} \mathrm{P}(x)dx = \int_{x_2}^{\infty} \mathrm{P}(x)dx = \delta. \tag{2.53}$$

Neglecting a part of the left tail should remove the divergence in the integrand at $x = -1/h$. The right tail is truncated accordingly to make the utility function 'fair'.

ii. Optimize the fraction f on the (x_1, x_2) domain

$$\max_f \int_{x_1}^{x_2} \mathrm{P}(x) \log(1 + fx)dx. \tag{2.54}$$

By choosing δ accordingly, we try to solve the trade-off of neglecting areas under $\mathrm{P}(x)$ (choose δ small enough) and having $x_1 > -1/f^*$ for the optimal fraction (choose δ large enough). The optimal fraction is derived by differentiating the integral with respect to fraction f:

$$\int_{x_1}^{x_2} \mathrm{P}(r) \left\{ \frac{\partial \log(1 + fx)}{\partial f} \right\}_{f^*} dx = \int_{x_1}^{x_2} \mathrm{P}(r) \left\{ \frac{x}{1 + f^* x} \right\} dx = 0 \tag{2.55}$$

The factor $\frac{x}{1+f^*x}$ of the integrand on (x_1, x_2) can be approximated by a Taylor Expansion of second order, $1 - f^*x$. According to Osorio (2008) this approximation holds for most practical applications including the case of the Student-T distribution with three degrees of freedom, having infinite kurtosis.
If we assume further that the neglected areas are small,

$$\begin{aligned} \left| \int_{-\infty}^{x_1} \mathrm{P}(r)x(1 - fx)dx \right| &<< 1 \\ \left| \int_{x_2}^{\infty} \mathrm{P}(r)x(1 - fx)dx \right| &<< 1 \end{aligned} \tag{2.56}$$

and the functions $\mathrm{P}(x)x(1-hr)$ and $\mathrm{P}(x)x/(1+hr)$ are close to each other between x_1 and x_2

$$\int_{x_1}^{x_2} \mathrm{P}(x)x(1-f^*x)dx \approx \int_{x_1}^{x_2} \mathrm{P}(x)\frac{x}{1+f^*x}dx. \tag{2.57}$$

The original derivative can be approximated by

$$\int_{\infty}^{-\infty} \mathrm{P}(x)x(1-f^*x)dx \approx 0. \tag{2.58}$$

Thus, the solution for the optimal Kelly fraction can be formulated as

$$f^* = \frac{\langle x \rangle}{\langle x^2 \rangle} = \frac{\mu}{\mu+\sigma^2}, \tag{2.59}$$

where $\langle \ldots \rangle$ denotes the average for the probability density function with mean μ and variance σ^2. This result holds for all probability density functions decaying fast enough, so that the probability mass below x_1 and above x_2 is neglectable and $f^* < 1/|x_1|$ to avoid singularity in the integrand (Osorio, 2008). Empirically, $\mu << \sigma^2$, simplifying the given solution to the already given solution $f^* = \mu/\sigma^2$ (Thorp, 2006). For the Student-T distribution Osorio shows that the approximation, using the second order Taylor approximation, breaks down at $\nu < 4$ as kurtosis tends to infinity.

2.7 Ornstein-Uhlenbeck processes (Lv and Meister, 2009)

Lv and Meister (2009) calculate the optimal investment strategy for financial assets following multiple Ornstein-Uhlenbeck processes. They state that their task is to "find the optimal self-financing trading strategy". For a complete and frictionless market, upon probability space on a time interval from 0 to T

$$(\Omega, \mathscr{F}_T, \mathrm{P}_T), \tag{2.60}$$

all random variables are specified. Ω denotes the sample space, $\mathscr{F}_T, t \in [0,T]$ is a σ-Algebra, capturing all available information up to time T and P_T is the according probability measure. Accordingly, the sub-probability space $(\Omega, \mathscr{F}_t, \mathrm{P}_t)$ is introduced.

Comparable to Merton (1992) in the GBM-framework, the risk-free asset follows the dynamics

$$dB_t = B_t r_t dt, \tag{2.61}$$

where r_t denotes the risk-free rate. The other assets $S_j(t), t \in [0,T], j \in [1,k]$ combine to the k-dimensional vector $S_t = (S_1(t), \ldots, S_k(t))^\top$. With the absolute number of assets in the risk free asset ϕ_0 and the risky assets $\phi_t = (\phi_1(t), \ldots, \phi_k(t))^\top$, the wealth of the portfolio can be stated as

$$W_t(\psi) = \phi_0 B_t + \phi_t S_t, \tag{2.62}$$

where $\psi = (\phi_0(t), \phi_t)$ is the set representing the trading strategy. The strategy will be called self-financing if

$$dW_t = \phi_0 dB_t + \phi_t dS_t, \ \forall t \in [0,T]. \tag{2.63}$$

Additionally the strategy will be called admissible iff

$$W_t(\psi) \geq 0, \ P_T \text{ a.s. } \forall t \in [0,T]. \tag{2.64}$$

For the class of admissible, self-financing trading strategies $\psi \in \mathscr{D}$, ψ^* is the optimal trading strategy iff

$$\mathsf{E}\left[U\left\{W_T\left(\psi^*\right)\right\}\right] \geq \mathsf{E}\left[U\left\{W_T\left(\psi\right)\right\}\right], \ \forall \psi \in \mathscr{D}, \tag{2.65}$$

whereas $U(x)$ is the utility function, concave in wealth.

Letting $r_t = r$ and $S_t(t) = \exp\left\{x_i(t)\right\}, t \in [0,T], j = 1, \ldots, k$ the process for the log-returns is

$$dx_t(t) = \left\{a_j - b_j x_j(t)\right\} dt + \sum_{l=1}^{k} \sigma_{j,l} dW_t^l, \ j,l = 1, \ldots, k, \tag{2.66}$$

where $a_j, b_j > 0$ and $\sigma_{j,l}$ are constants and W_t is a standard Brownian motion. In matrix notation the dynamics of the log-returns can be rewritten as

$$dx_t = (a - bx_t)\, dt + \sigma dW_t, \tag{2.67}$$

with $a = (a_1, \ldots, a_k)^\top$, $k \times k$ matrix b with $(b_1, \ldots, b_k)$ in the diagonal, else zero and σ as matrix from the Cholesky-factorization of the covariance matrix. Following Ito's Lemma, the dynamics of the prices $S_t = (S_1(t), \ldots, S_k(t))^\top$ are given as

$$dS_j(t) = S_j(t)\mu_j(t)dt + S_j(t) \sum_{l=1}^{k} \sigma_{j,l} dW_t^l, \;\; j, l = 1, \ldots, k, \tag{2.68}$$

where $\mu_j(t) = a_j - b_j \log(S_j(t)) + \frac{1}{2}\|\sigma_j\|^2$ and $\sigma_j = (\sigma_{j,1}, \ldots, \sigma_{j,k})^\top$.

Theorem 2.7.1 (Lv and Meister)**.** *The optimal fraction vector* $f_t^* = (f_1^*(t), \ldots, f_k^*(t))^\top$ *is rearranged from* $f_j^*(t) = \frac{\phi_j^*(t) S_j^*(t)}{W_t^*}$ *to*

$$f_t^* = \Sigma^{-1} \left\{ \mu_t(t) - r \right\}, \tag{2.69}$$

where $\mu_t(t) = a - b \log(S_t(t)) + \frac{1}{2}\|\sigma_j\|^2$. $\Sigma = \sigma\sigma^\top$ *is the covariance matrix in order to model the correlated Brownian motions in the Ornstein-Uhlenbeck processes for the risky assets.*

In conclusion, the approach of Lv and Meister (2009) is compatible with the continuous time framework of Merton (1992). Solely the drift structure for the fixed-income instruments needs to be modified in order to make them applicable.[2]

2.8 Hierarchical Bayes Model (MacLean,Ziemba and Li, 2005)

MacLean and Ziemba (2005) and MacLean et al. (2005) derive the solution of the

[2]If all assets would be perfectly correlated plus having the same mean, the optimal fraction would reduce to $f^*(t) = \frac{\mu_1(t) - r}{\sigma^2}$, the univariate solution of the optimal fraction for the Kelly Criterion.

Kelly growth-optimum strategy in a hierarchical Bayesian framework. Given asset prices $P(t) = (P_0(t), \ldots, P_k(t))^\top$ for assets $j = [0, \ldots, k]$, discrete returns over a period from t to T are given as $R_j(t,T) = 1 + r_j(t,T),\ j = [0, \ldots, k]$. The investment strategy in t is formulated as $f(t) = (f_0(t), \ldots, f_k(t))^\top$. Accordingly, the wealth in point T is given as

$$W_T = W_t R^\top(t,T) f(t) \tag{2.70}$$

$$= W_t \sum_{j=0}^{k} R_j(t,T) f_j(t). \tag{2.71}$$

Partitioning the time interval from t to T into n parts with $d = \frac{T-t}{n}$, so that the investment strategy may be re-balanced at $t, t+d, \ldots, t+nd = T$, the wealth process changes to

$$W_T = W_t \prod_{i=0}^{n-1} R^\top \left\{t + id, t + (i+1)d\right\} f(t + id) \tag{2.72}$$

$$= W_t \left[\exp \left\{ \frac{1}{n} \sum_{i=0}^{n-1} \log R^\top (t + id, t + (i+1)d) f(t + id) \right\} \right]^n. \tag{2.73}$$

The exponential form emphasizes the growth rate with strategy $f(t)$:

$$G_n(f) = \frac{1}{n} \sum_{i=0}^{n-1} \log \left[R^\top \left\{t + id, t + (i+1)d\right\} \right] f(t + id). \tag{2.74}$$

As $d \to 0$, the wealth process becomes continuous and for $d \to \infty$, W_T converges to a log-normal random variable.[3] The goal of the Kelly growth-optimum strategy is to maximize the geometric mean of the discrete wealth process or alternatively, maximize the mean of the logarithm of wealth.[4]

[3] Note, if r.v. x is normally distributed, $\exp(x)$ is log-normally distributed.

[4] $\log x_{\text{geom}} = \frac{1}{n} \sum_{i=1}^{n} \log x_i$

2.8.1 The pricing model

The pricing model itself is formulated in continuous time. The infinitesimal increment process for the assets are defined as stochastic linear equations:

$$d\log P_0(t) = r dt, \tag{2.75}$$

$$d\log P_j(t) = \alpha_j dt + \delta_j dV_j,\ j = 1, \ldots, k, \tag{2.76}$$

with r as risk free rate. $V_j,\ j = 1, \ldots, k$ are independent Brownian motions and the mean parameters are random variables itself:

$$d\alpha_j(t) = \mu_j dt + \gamma_j dZ_j,\ j = 1, \ldots, k. \tag{2.77}$$

The Brownian motions Z_j are correlated with $\rho_{i,j}$ for Z_i and Z_j. The correlation in the presented model arise solely from the correlations generated by the factors in the expected returns. Stochastic volatility factors for the correlation are not captured. MacLean and Ziemba (2005) note that the hierarchical linear model is a generalization of the single stock model in Browne and Whitt (1996). The mean rates α_j are stochastic and the volatilities δ_j are not stochastic. This is due to the main result of Chopra and Ziemba (1993), that errors in the mean estimate are most crucial in order to estimate the portfolio fraction. Hence, for $y(t) = \log\{P_j(t + id)\}$

$$y(t) = [y_1(t), \ldots, y_k(t)]^\top, \tag{2.78}$$

$$\alpha = [\alpha_1, \ldots, \alpha_k]^\top, \tag{2.79}$$

$$\Delta = [\delta_1^2, \ldots, \delta_k^2]^\top, \tag{2.80}$$

$$\mu = [\mu_1, \ldots, \mu_k]^\top, \tag{2.81}$$

$$\Gamma = (\gamma_{i,j}) = (\gamma_i \gamma_j \rho_{i,j}), \tag{2.82}$$

representing log-prices, drift rate, variances, expected drift rate and covariances from the factors in the expected returns. Under log-normality of the price-process

given (α, Δ),

$$(y(t) \mid \alpha, \Delta) \propto \mathrm{N}(\mu \cdot t, t \cdot \Delta). \tag{2.83}$$

The prior distribution of the drift rate from formula 2.77 is given by

$$\alpha \propto \mathrm{N}(\mu, \Gamma). \tag{2.84}$$

The marginal is given as

$$y(t) \propto \mathrm{N}(\mu(t), \Sigma(t)), \tag{2.85}$$

with $\Sigma(t) = t^2\Gamma + t\Delta = \Gamma(t) + \Delta(t)$. The covariance for log-prices is partitioned into a component determined by the random drift and a component determined by the diffusion (MacLean and Ziemba, 2005).

2.8.2 Bayes estimation

Assuming that in each point in time t, $\{y(s),\ 0 \leq s \leq t\}$, historical data are available with according filtration $\mathscr{F}_t^y = \sigma\{y(s),\ 0 \leq s \leq t\}$. The realized mean rate of return is

$$\bar{y}_t = \frac{1}{t} y(t). \tag{2.86}$$

With the formulated prior distribution in Formula 2.77, the posterior distribution for the conditional mean rate of return is

$$(\alpha \mid \mathscr{F}_t^y) \propto \mathrm{N}(\hat{\alpha}_t, \hat{\Gamma}_t), \tag{2.87}$$

with

$$\hat{\alpha}_t = \mathsf{E}(\alpha \mid \mathscr{F}_t^y) = \mu(t) + \left\{I - \Delta(t)\Sigma^{-1}(t)\right\}\left\{\bar{y}(t) - \mu(t)\right\}, \tag{2.88}$$

$$\hat{\Gamma}_t = \frac{1}{t^2}\left\{I - \Delta(t)\Sigma^{-1}(t)\right\}\Delta(t). \tag{2.89}$$

The parameter set (α, Γ, Δ) needs to be estimated from data in order to provide an empirical Bayes estimate for the conditional mean rate of return.

The covariance matrix $\hat{\Sigma}_t$ is estimated by using the spectral decomposition:

$$\hat{\Sigma}_t = \Upsilon \Lambda \Upsilon^\top, \tag{2.90}$$

with Υ, the a diagonal matrix of eigenvectors and Λ, a diagonal matrix with the eigenvalues ψ_j in the diagonal. The matrix Λ is truncated by setting, for a pre-specified value $m, m < k$, $\psi_{m+1}, \ldots, \psi_k$ to zero. The resulting matrix is

$$G_t = \Upsilon \Lambda_{trunc} \Upsilon^\top. \tag{2.91}$$

This is the estimate for $\hat{\Gamma}_t$. Analogously,

$$D_t = diag(\Sigma_t - G_t), \tag{2.92}$$

with $\frac{t}{n} D_t$ as estimator for $\hat{\Delta}_t$. Again, $\hat{\Sigma}_t = \hat{\Delta}_t + \hat{\Gamma}_t$.
The parameter μ is estimated by the common mean over all log-returns, implying

$$\hat{\mu}_t = \left\{ \frac{1}{Tk} \sum_{j}^{k} \sum_{t}^{T} y_j(t) - y_j(t-1) \right\} \times \mathbf{1}. \tag{2.93}$$

Hence, the truncation estimator for the conditional mean rate of return at time t is, according to MacLean and Ziemba (2005) given by

$$\hat{\alpha}_t = \mathsf{E}(\hat{\alpha} \mid \mathscr{F}_t^y) = \hat{\mu} + (I - \hat{\Delta}_t \hat{\Sigma}_t^{-1})(\bar{y}_t - \mu). \tag{2.94}$$

MacLean et al. (2005) point out that if the model is correct, $\hat{\alpha}_t$ induces a smaller mean square error than maximum-likelihood estimates or James-Stein estimates.

2.8.3 The Optimization

The discrete optimization goal of the Kelly growth-optimum strategy is to maximize the expectation of formula 2.74, the growth rate of wealth. The average growth rate is therefore given as

$$\mathsf{E}\left\{G_n(f)\right\} = \frac{1}{n} \sum_{i=0}^{n-1} \mathsf{E}\left\{\log R^\top(t+id, t+(i+1)d) f(t+id)\right\}. \tag{2.95}$$

Given serially independent log-returns, the maximization of $E\{G_n(f)\}$ is

$$\max_f \left[\mathsf{E}\left\{\log(R^\top(t+id, t+(i+1)d)f(t+id)\right\}\right], \tag{2.96}$$

for each i. If the distribution is the same for each i, the strategy will be fixed over time. For the hierarchical Bayesian case under the normal, the goal is to

$$\max_f \left\{(\varphi - r\mathbf{1})^\top + r - \frac{1}{2}f^\top \Sigma f\right\}, \tag{2.97}$$

with $\varphi = \alpha_j + \frac{1}{2}\delta_j^2$. For the continuous time problem, in accordance with Merton (1992), the optimal investment fraction is given as

$$\tilde{f}^* = \Sigma^{-1}(\varphi - r\mathbf{1}), \tag{2.98}$$

where $f^* = (f_0, \tilde{f}^*)$, with f_0 as the risk-free investment fraction defined as $f_0 = 1 - \sum_{j=1}^k f_j$. The strategy is an approximation of the discrete investment optimization problem and is seen as fixed-mix strategy, as the the investment portfolio needs to be re-balanced with varying wealth in order to maintain the fractions.

Chapter 3

Risk-Constrained Kelly portfolios under alpha-stable laws

3.1 Introduction

Given a set of investment opportunities, how should the investment weights be chosen in order to have more wealth than any other investor at the end of the investment period? The Kelly growth-optimum strategy is a betting scheme for an investor, who seeks to asymptotically maximize his growth rate of capital. This strategy outperforms any other significantly different strategy, given knowledge of the true underlying process (Breiman, 1961). But, the sole use of the Kelly Criterion implies larger bets than a representative, risk-averse investor would accept in terms of risk (Hausch and Ziemba, 1985; Clark and Ziemba, 1987). Thus, the Kelly optimization needs to be restricted by a risk measure. We use α-stable laws and its scaling behavior in order to model the underlying financial market returns. Upon the Generalized Central Limit Theorem (GCLT), the horizon distribution is modelled in an discrete i.i.d. framework.

The aim is to maximize the geometric portfolio return, i.e. Kelly Criterion and

restrict the objective to a subjective risk constraint, formulated as spectral risk measure, including quantile (VaR) or Expected Shortfall as special cases. The formulated trade-off introduces a mapping over growth and risk in order to evaluate the investment decision. The contribution of this paper is three-fold: The first contribution represents the application of multidimensional α-stable laws, in the form of elliptically α-stable distributions, to the constrained Kelly portfolio. Second, instead of simulating from the class of elliptically α-stable distributions, a semiparametric scaling approximation, based on the data set itself, is proposed. Third, assets with non-linear payoff structure, long put-options, are incorporated into the nonlinear optimization to allow for asymmetric payoffs, which lead to a higher growth criterion, given a fixed risk constraint.

The Kelly Criterion originates from Kelly (1956), dealing with, from the point of information theory, an optimal investment strategy in a binary channel. Breiman (1961) formally proves the asymptotic outperformance of the Kelly strategy for arbitrary distributions in an i.i.d. world. For arbitrarily distributed, possibly non-stationary processes, those results have been extended by Algeot and Cover (1988). Incorporating risk measures into the Kelly optimization, MacLean et al. (1992) discuss the growth-risk trade-off in terms of efficiency. Roll (1973) compares the Markowitz arithmetic mean maximization with the Kelly geometric mean maximization. In contrast to Constant Proportion Portfolio Insurance (CPPI), the investment strategy remains fixed fraction, given stationarity. More recently, Busseti et al. (2016) introduce an alternative risk constraint, limiting the probability of a drawdown of wealth to a given undesirable level.

The distribution of financial market returns for a chosen horizon is modelled as the sum of hourly random variables. As the distribution in some horizon is presumed to be non-Gaussian, the classical Central Limit Theorem (CLT) does not apply as second and higher moments may not exist. Thus, the generalized central limit theorem (GCLT) of Gnedenko and Kolmogorov (1954b) is applied for the sum of

random variables, whose second and higher moments may not be bounded. For the financial application this implies the use of α-stable laws (Lévy, 1925; Mandelbrot, 1963b; Fama, 1965). As multidimensional α-stable random variables are difficult to evaluate for larger dimensions, elliptical α-stable distributions are employed, allowing for efficient portfolio estimation for dimensions $k \leq 40$ (Nolan, 2013) in the presence of linear dependence.

Price data, both for assets with linear and non-linear payoff structure, were gathered from Lobster and Bloomberg. For computation, Matlab 2016a was utilized. In order to solve the formulated nonlinear optimization problem the sequential quadratic algorithm in *fmincon* was employed.

The paper is organized as follows: In Chapter one the portfolio allocation problem is stated. The financial model is formulated by using generalized measures for growth and risk. Chapter two, the estimation, starts with a case for non-Gaussianity of financial log-returns of different sampling frequencies, reasoning the utilization of α-stable laws. For the multidimensional case, elliptically α-stable distributions are introduced in order to have an analytically tractable class of distributions. As the semiparametric scaling approximation is introduced, the estimation of location and scale is illustrated. An application is given in Chapter three, the implementation. For a representative investor with a planning horizon of one year, the optimally VaR/ES-constrained Kelly portfolios are found, benefitting from the protective put strategy.

3.2 Model

3.2.1 Portfolio allocation problem

Given initial wealth of the investor $W_0 \in \mathbb{R}^+$, there are $j = 1, ..., k$ investment opportunities with fractions $f_t = [f_{1,t}, \ldots, f_{k,t}]^\top \in \mathbb{R}^k$ in period $t = 1, 2, \ldots, T$.

$T \in \mathbb{N}^+$ represents the planning horizon. Assessing solely self-financing strategies, the budget constraint is given by $\sum_{j=1}^{k} f_{j,t} \leq 1$. Given a statistical model for continuous returns $X_t \in \mathbb{R}^k$, discrete returns are calculated by $\tilde{X}_t = \exp\{X_t\} - 1$. Given outcomes in $t-1, \ldots, T$ the wealth in T is given by

$$\begin{aligned} W_T(f_t) &= W_0 \prod_{t=1}^{T} \left\{ 1 + \sum_{j=1}^{k} f_{j,t} \tilde{X}_{j,t} \right\} \\ &= W_0 \prod_{t=1}^{T} \left\{ 1 + f_t^\top \tilde{X}_t \right\}. \end{aligned} \tag{3.1}$$

Given the stochastic wealth process, measures for growth and risk are formulated in order to choose investment fractions f_t, which suit investor preferences.

For a cdf $F_{W_T}(x)$ the spectral risk/growth measure with weight function $\phi(x)$ is defined through the quantile function $F^{-1}_{W_T}(x) \stackrel{\text{def}}{=} \{x : \mathrm{P}(W_T(f_t) \leq x) = \alpha\}$, $\alpha \in (0, 1)$.

$$M_\phi \{W_T(f_t)\} = \int_0^1 \phi(x) F^{-1}_{W_T}(x) \mathrm{d}x \tag{3.2}$$

Within the context of spectral risk measures, the measure will be coherent iff the weight function is positive $\phi(x) \geq 0$, increasing $\phi'(x) \geq 0$ and normalized $\int_0^1 \phi(x) = 1$ (Acerbi, 2002). For the discrete framework (3.1) with $n \in \mathbb{N}^+$ wealth trajectories, the measure is defined as

$$M_\phi \{W_T(f_t)\} = \sum_{i=1}^{n} \phi_i W_{T,i}(f_t), \tag{3.3}$$

where $W_{T,i}$ denotes element i out of n wealth paths with weight ϕ_i.

Growth measures

Following Roll (1973), there are two main strands dealing with the accumulation of wealth and thus, the allocation of wealth into a portfolio. On the one

hand, the Markowitz optimization aims to maximize the expected portfolio return (Markowitz, 1952; Tobin, 1958; Sharpe, 1964; Lintner, 1965). On the other hand, the Kelly growth-optimum approach by Kelly (1956), Breiman (1961) and Thorp (1971), aims to maximize the expected logarithm of wealth, which is equivalent to maximizing the geometric portfolio return. Within the framework of spectral growth/risk measures, the growth measures for the Markowitz and the Kelly optimization are evaluated:

- G_1 : For the expected wealth, the growth criterion from the Markowitz optimization, the weight function is

$$\phi_E(x) = 1,$$

giving

$$G_{\phi_E}\left\{W_T(f_t)\right\} = \int_0^1 F_{W_T}^{-1}(x)\mathrm{d}x = \mathsf{E}\left\{W_T(f_t)\right\}. \tag{3.4}$$

- G_2 : The expected logarithm of wealth, representing the optimization criterion for the Kelly strategy, is obtained for the weight function

$$\phi_{\mathrm{Elog}}(x) = \log(x),$$

giving

$$G_{\phi_{\mathrm{Elog}}}\left\{W_T(f_t)\right\} = \int_0^1 \log F_{W_T}^{-1}(x)\mathrm{d}x = \mathsf{E}\left\{\log W_T(f_t)\right\}. \tag{3.5}$$

The growth measure will be denoted by $G_\phi\left\{W_T(f_t)\right\}$ and the optimization for horizon T without risk constraints is formulated as

$$\max_{f_t\in\mathbb{R}^k}\left[G_\phi\left\{W_T(f_t)\right\} \,\middle|\, \sum_{j=1}^{k} f_{j,t} \leq 1\right]. \tag{3.6}$$

This paper focusses on the Kelly growth criterion as it represents a betting scheme for an investor, who seeks to asymptotically maximize his growth rate of capital. The betting strategy outperforms any other significantly different strategy

asymptotically and minimizes the expected time to reach a goal (Breiman, 1961; Algeot and Cover, 1988). For a comprehensive treatment of the Kelly Criterion, see MacLean et al. (2011). Whereas the maximization of the expected wealth in the Markowitz optimization, given favorable investment possibilities, always implies betting the entire fortune, the maximization of the expected logarithm of wealth leads to one growth-optimal portfolio, which is not necessarily optimal in terms of the Markowitz portfolio (Thorp, 1971). Accordingly Markowitz (1976) considers the Kelly portfolio to be the upper limit for a conservative investor. Furthermore, the log-optimal strategy is fixed fraction, independent of time (MacLean et al., 1992).

Risk measures

The sole use of the Kelly Criterion implies larger bets than a representative, risk-averse investor would accept in terms of risk (Hausch and Ziemba, 1985; Clark and Ziemba, 1987). In order to formulate individual risk measures for different investors, the spectral risk measure from (3.2), denoted by $S_\phi\{W_T(f_t)\}$, will be used. Two specific risk measures to include the degree of risk-aversion into the portfolio optimization are quantile (Value at Risk) and conditional tail expectation (Expected Shortfall) constraints:

- S_1 : The quantile constraint (VaR) is a special case of the spectral risk measure from (3.2)

$$\phi_{\mathrm{Q}_\alpha}(x) = \delta(x = \alpha),\ \alpha \in (0, 1), \tag{3.7}$$

  where $\delta(x = \alpha)$ is the Dirac delta function, well known to be a non-coherent risk measure. Further drawbacks of the quantile constraint are treated in Basak and Shapiro (2001). However, the quantile restriction allows to ask

the investor specifically to name a fraction of his wealth he can accept to lose with probability $1 - \alpha$.

- S_2 : In contrast, Conditional Tail Expectation (ES) is a coherent risk measure representing the average loss beyond a given quantile constraint. Being a special case of the spectral measure, the weight function is given as

$$\phi_{\mathrm{CTE}_\alpha}(x) = \alpha^{-1}\mathbf{1}(x < \alpha). \tag{3.8}$$

Growth-risk frontier

Following MacLean et al. (1992), the possible combinations of growth and risk measures are given by the set

$$U = \left[G_\phi\left\{W_T(f_t)\right\}, S_\phi\left\{W_T(f_t)\right\}\right], f_t \text{ feasible.} \tag{3.9}$$

The growth-risk frontier is accordingly formulated as

$$U_t^* = \left[G_\phi\left\{W_T(f_t^*)\right\}, S_\phi\left\{W_T(f_t^*)\right\}\right], f_t^* \text{ feasible,} \tag{3.10}$$

where the $f_t^* \in \mathbb{R}^k$ is the investment fraction maximizing the growth measure under risk restriction.

$$\begin{aligned} f_t^* = \underset{f_t^* \in \mathbb{R}^k}{\arg\max}\ & G_\phi\left\{W_T(f_t)\right\} \\ \text{s.t. } & S_\phi\left\{W_T(f_t)\right\} \leq b,\ b \in \mathbb{R}, \\ & \sum_{j=1}^{k} f_{j,t} \leq 1 \end{aligned} \tag{3.11}$$

For the Kelly Criterion with a risk constraint as proposed, the frontier is illustratively visualized in Figure 3.1. In contrast to the Markowitz maximization, implying a steady tradeoff between mean and risk, the geometric mean maximization implies one specific portfolio - the Kelly portfolio - exhibiting the highest

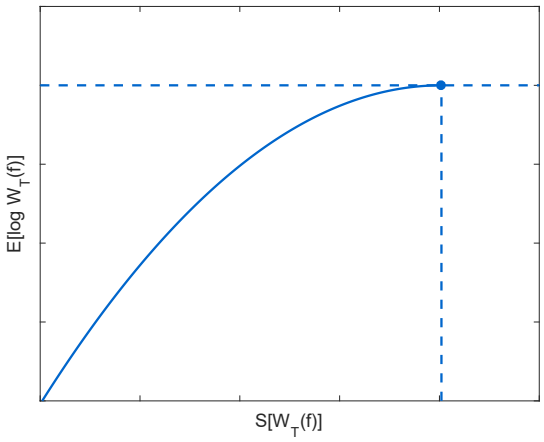

Figure 3.1: Kelly-risk frontier with unconstrained Kelly portfolio exhibiting the highest geometrical mean

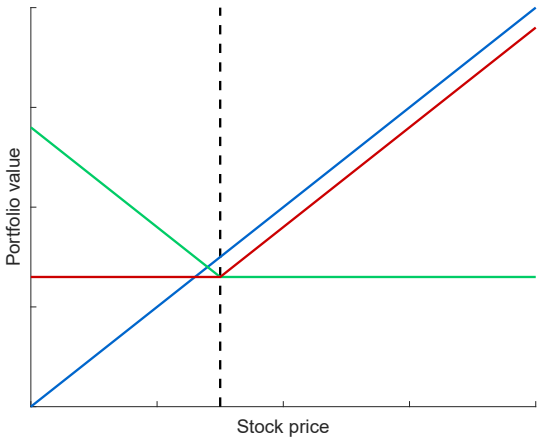

Figure 3.2: Protective put strategy (red) consisting of long stock (blue) and long put (green) with chosen strike (dotted black)

geometric mean possible (horizontal dotted line). From this viewpoint, portfolios exhibiting a larger risk constraint than the Kelly portfolio (to the right of the vertical dotted line) are not efficient. If the investor prefers a smaller risk constraint than the full Kelly investor, restricted Kelly portfolios (solid line) constitute the Kelly-risk frontier. These are portfolio strategies with the highest growth criterion given risk constraint.

3.2.2 Tail constraints and non-linear instruments

The introduction of assets as nonlinear functions of the underlyings, derivatives, allows for controlling the asymmetry of the wealth distribution in such a way, that it will be skewed to the left. Albeit the distribution of the risk measure, the loss of the portfolio is limited by construction for high confidence levels. The instruments to achieve the asymmetric payoff profile are long put options. By construction, corridor options, as argued in the context of quantile constraints, are circumvented (Basak and Shapiro, 2001). A simplified representation of the

protective put strategy is given in Figure 3.2, consisting of one stock (blue) and one long put option (green) with chosen strike (dotted black). The result is the protective put strategy (red). The difference in payoff above the strike level is due to the put price, which the option holder has to pay. For $k \in \mathbb{N}^+$ linear assets with multiple put options each, given a pre-specified horizon, the choice of the fraction of linear and nonlinear assets is not obvious.

3.3 Estimation

3.3.1 A case for non-Gaussianity

Although Fama (1965) finds evidence for α-stable characteristics for all returns of the Dow Jones Index, it can be observed that financial (log-)returns tend to the Gaussian distribution as the sampling frequency decreases, see also McFarland et al. (1982), Boothe and Glassman (1987), and Dacorogna et al. (2001). The subsequent textbook example for the Standard and Poor's 500 reads as Table 3.1. Due to the 2009 financial crisis, an outlier week of -60% increases (decreases) the sample kurtosis (skewness) for the weekly frequency significantly from 13.67 (-1.27) to 131.09 (-6.7). If the outlier week is omitted, see column S&P (weekly*) of table 3.1, the general observation of decreasing kurtosis and increasing negative skewness is supported for *different* sample sizes. Still, including the outlier week of 2009, erratic behavior of sample moments definitely appears for this reference data series.

Table 3.1 Log-return descriptives for different sampling frequencies, S&P 500 1985-2015, frequency weekly* omits one week in the financial crisis 2009

Descriptives	S&P (daily)	S&P (weekly)	S&P (weekly*)	S&P (monthly)	S&P (yearly)
Data points	7564	1513	1512	360	30
Mean (p.a. in %)	8.37	8.37	10.41	8.37	8.37
Std (p.a. in %)	18.35	20.93	17.96	16	16.61
Skewness	−1.29	−6.7	−1.27	−1.98	−1.78
Kurtosis	31.26	131.09	13.67	12.48	7.13

The empirical observation of Gaussian convergence for lowering sampling frequencies cannot be shown explicitly by existing data, as data-records capture only 7564 trading days, representing 30 years of data. The empirical verification would require an appropriately large number of weeks, months and years.

In order to show that the annual return distribution, consisting of 30 data points, is with large probability not Gaussian, we randomly sample 10^5 30 blocks of daily returns from the S&P 500 and calculate second, third and fourth moments in order to evaluate dispersion, skewness and leptokurtic behavior (Bootstrap). Hence, for the three moments, the block-bootstrap estimators are plotted as histogram in Figure 3.3. The vertical red lines represent the moment estimators for the whole daily data series. In essence, dispersion, skewness and especially leptokurtic behavior of the bootstrap estimators are significantly biased, compared to the estimator of the whole series. Fewer sampled data-points imply less probability of sampling data in the tails of the return distribution. There was not one out of 10^5 30 day sub-samples, which resulted in a comparable kurtosis of the complete data-series. The result holds for sampling 30 separate days randomly under the i.i.d. assumption. Moments of order larger than one behave erratically over an increasing data sample, as first analyzed for commodity prices in Mandelbrot (1963b). Figure

3.4 plots standard deviation (in %), skewness and kurtosis as function of the used data-points of the series. The red lines represents the empirical moment behavior with increasing daily data points. The blue lines represent 100 trajectories of Gaussian moments with increasing data points. The observation of erratic moment behavior stands in contrast to Gaussian behavior. The observation holds over sampling frequencies daily, weekly, monthly and annually. This specific sample-size problem is crucial in risk management, especially for estimating quantiles of high confidence of the wealth distribution as in the constrained portfolio optimization in Equation 3.11. As the confidence level tends to one, having only a limited amount of data, the quantile estimate is systematically biased as the quantile is overestimated. The portfolio analyst has to evaluate if the estimated quantile given the chosen confidence level still has an acceptable distribution.

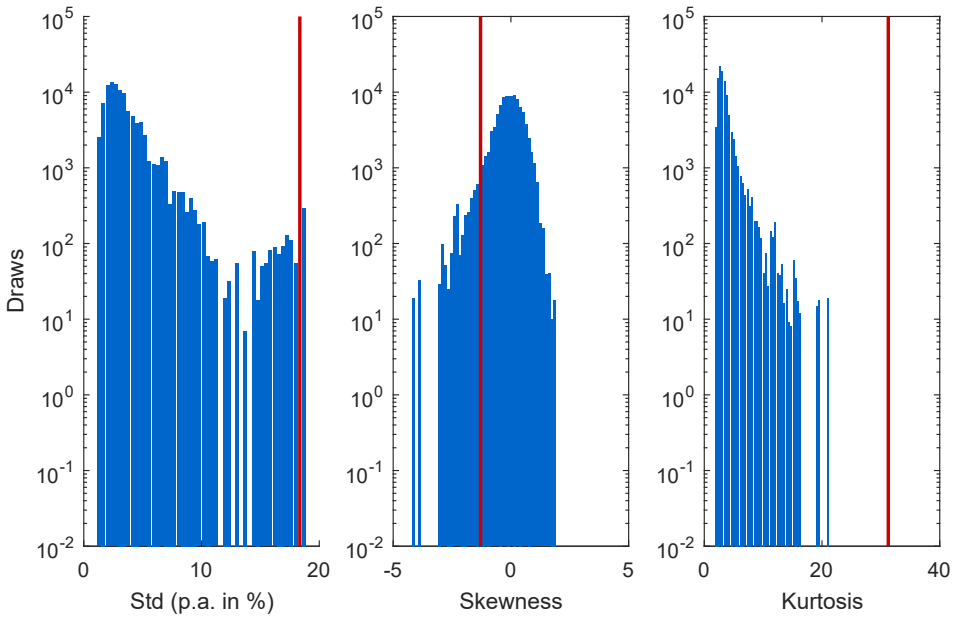

Figure 3.3: Whole sample (red) and block-bootstrapped standard deviations (p.a.), skewness and kurtosis for 10^5 draws of 30 subsequent daily returns (blue) from the S&P 500, 1985 to 2015

Stable_Kelly_MomentDistribution

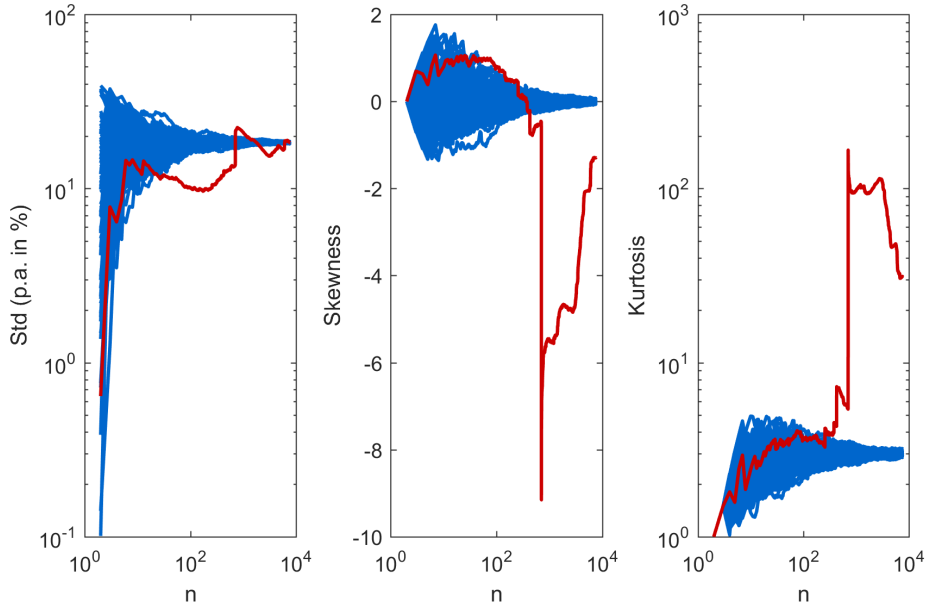

Figure 3.4: (Log-)Log plots of standard deviation (in %), skewness and kurtosis with increasing data points, S&P 500 from 1985 to 2015 (red) and 100 Gaussian simulations with S&P 500 moments (blue)

Stable_Kelly_MomentIncrease

Consequently, for investors with longer investment horizons, such as a year, the sum of daily random variables, constituting the yearly distribution, should not converge to the Gaussian, but to a heavy-tailed distribution, which will turn out to be the class of α-stable distributions. For financial markets, this assumption will imply infinite variance, skewness and kurtosis, leading to non-converging moments, i.e. the observed erratic behavior. The model of Chapter 3.2 will be estimated within a stationary framework for elliptically α-stable distributions, striving for scale invariance. Although daily and higher frequency returns exhibit non-stationary characteristics, the horizon distribution, i.e. yearly, cannot be shown to exhibit significant volatility clustering.

3.3.2 Scale invariance

Let $X_t \in \mathbb{R}^k$ be a multidimensional, i.i.d. random variable from distribution P^t, where t indicates the scale e.g. days. Given the investment horizon of the investor, T days, the wealth Equation 3.1

$$W_T(f_t) = W_0 \prod_{t=1}^{T} \left\{1 + f_t^\top \tilde{X}_t\right\} = W_0 \prod_{t=1}^{T} \left\{f_t^\top \exp(X_t)\right\} \tag{3.12}$$

can be simplified, given $f_t = f \ \forall t = 0, \ldots, T$.

$$W_T(f) = W_0 \left\{ f^\top \exp\left(\sum_{t=1}^{T} X_t\right)\right\} = W_0 \left\{f^\top \exp(X)\right\}, \ X \stackrel{\text{def}}{=} \sum_{t=1}^{T} X_t \tag{3.13}$$

As the horizon T grows, the sum of the random variables X_t tends to the Gaussian as long as the first two moments of the underlying distribution are finite. Formally, let random variable X_t have expectation vector $\mu_t = \mathsf{E}(X_t)$ and covariance matrix $\sigma_t = \mathsf{E}\left[\{X_t - \mathsf{E}(X_t)\}\{X_t - \mathsf{E}(X_t)\}^\top\right]$. Then

$$\begin{aligned} \sum_{t=1}^{T} X_t &\xrightarrow{\mathcal{L}} \mathrm{N}\left(\sum_{t=1}^{T} \mu_t, \sum_{t=1}^{T} \sigma_t\right) \\ X &\xrightarrow{\mathcal{L}} \mathrm{N}(\mu, \sigma) \\ T^{-\frac{1}{2}} \sum_{t=1}^{T} (X_t - \mu_t) &\xrightarrow{\mathcal{L}} \mathrm{N}(0, \sigma). \end{aligned} \tag{3.14}$$

If the distribution in horizon T is modelled as the sum of higher frequency distributions, the multidimensional process of returns, which may not be Gaussian, but of finite variance, converges to the Gaussian. In contrast, as argued in Section 3.3.1, returns of horizons beyond the sampled frequency, are presumed to be heavy-tailed. Hence, the standard Central Limit Theorem (CLT) does not apply. Except for the Gaussian itself, finite variance distributions change their shape under aggregation. In contrast, the class of α-stable distributions is scale invariant

(Mandelbrot, 1963b). Scale invariance of distribution P is defined via a continuous function g, such that for all x

$$g(\lambda)P(x) = P(\lambda x), \tag{3.15}$$

with $\lambda x \geq x_0$ and $x_0 > 0$. Equivalently, distribution P has a power-law tail, implying that for $x \geq x_0 \geq 0, c \geq 0$ and $\alpha > 0$

$$P(x) = cx^{-\alpha}. \tag{3.16}$$

In that respect, a one-dimensional random variable $X \sim \mathrm{S}(\alpha, \beta, \gamma, \delta)$ will be α-stable distributed with parameters $0 < \alpha \leq 2,\ -1 \leq \beta \leq 1,\ \gamma \geq 0$ and $\delta \in \mathbb{R}$ (Nolan, 2017; Cizek et al., 2011), if

$$X \overset{\mathcal{L}}{=} \begin{cases} \gamma Z + \delta, & \alpha \neq 1 \\ \gamma Z + (\delta + \beta \frac{2}{\pi} \gamma \log \gamma), & \alpha = 1. \end{cases} \tag{3.17}$$

$\mathrm{S}(Z \mid \alpha, \beta, 1, 0)$ represents the standard α-stable form. Only special cases of α-stable distribution are available as real-valued densities (e.g. Gaussian, Cauchy and Lévy).

Scale invariance under addition implies that for the sum of α-stable variables $X_t \sim \mathrm{S}(\alpha, \beta, \gamma, \delta_t),\ t = 1, \ldots, T$

$$X_1 + X_2 + \ldots + X_T = \sum_{t=1}^{T} X_t = X \sim \mathrm{S}(\alpha, \beta, T^{\frac{1}{\alpha}}\gamma, \delta), \tag{3.18}$$

where $\delta = T\delta_t$.

According to Gnedenko and Kolmogorov (1954b), the limiting distribution of T i.i.d. α-stable random variables, $0 < \alpha \leq 2$ is

$$a_T \left(\sum_{t=1}^{T} X_t \right) - b_T \xrightarrow{\mathcal{L}} \mathrm{S}(\alpha, \beta, 1, 0), \tag{3.19}$$

where $a_T > 0$ and $b_T \in \mathbb{R}$. The special case of the Generalized Central Limit Theorem (GCLT) is the CLT of Equation 3.14 for $\alpha = 2$, $\beta = 0$, $\gamma = \frac{\sigma}{\sqrt{2}}$ and

$\delta_t = \mu_t$, given $a_T = \frac{1}{\sigma\sqrt{T}}$ and $b_T = \frac{\sqrt{T}\mu}{\sigma}$. In general, for $0 < \alpha \leq 2$,

$$T^{-\frac{1}{\alpha}} \sum_{t=1}^{T} (X_t - \delta_t) \xrightarrow{\mathcal{L}} \mathrm{S}(\alpha, 0, \gamma, 0). \tag{3.20}$$

3.3.3 Elliptically contoured α-stable distributions

For the multidimensional estimation, α-stable laws are not extensively accessible as closed-form densities are only available for special cases. One computationally tractable exception are elliptically contoured α-stable laws, which can be efficiently estimated for dimensions $k \leq 40$ (Nolan, 2013). This class of distributions enables the modeling of heavy tails while preserving its shape under aggregation in the presence of linear dependence.

Random vector $Y = [Y_1, \ldots, Y_k]^\top$ has a spherical distribution iff the characteristic function $\varphi_Y(u)$ satisfies for all $u \in \mathbb{R}^k$

$$\varphi_Y(u) = \mathsf{E}\left\{\exp\left(iu^\top Y\right)\right\} = \psi(u^\top u) = \psi(u_1^2 + \ldots + u_k^2), \tag{3.21}$$

where ψ is the characteristic generator of the spherical distribution.

Random vector $X \sim \mathrm{S}_k(\delta, \Gamma, \psi)$ is elliptically distributed with positive definite scaling matrix $\Gamma = AA^\top$, $A \in \mathbb{R}^{k \times k}$ and location vector $\delta \in \mathbb{R}^k$ when

$$X \overset{\mathcal{L}}{=} \delta + AY, \tag{3.22}$$

where Y is spherical with characteristic generator ψ. The characteristic function is given by

$$\varphi_X(u) = \mathsf{E}\left\{\exp\left(iu^\top X\right)\right\} = \exp\left(iu^\top \delta\right) \psi\left(u^\top \Gamma u\right). \tag{3.23}$$

A subclass of elliptical distributions are normal variance mixtures

$X = [X_1, \ldots, X_k]^\top$ for

$$X \overset{\mathcal{L}}{=} W^{1/2} AZ + \delta, \tag{3.24}$$

with $Z \sim \mathrm{N}(0, I_k)$ and $W \geq 0$ being a non-negative one-dimensional random variable, independent of Z (Kring et al., 2009).

A further subclass of normal variance mixtures are α-stable sub-Gaussian $X = [X_1, \ldots, X_k]^\top$ for $W \sim \mathrm{S}(\alpha/2, (\cos \pi\alpha/4)^{2/\alpha}, 1, 0)$, $0 < \alpha < 2$, being α-stable distributed, parameterized following Nolan (2017). $G \sim \mathrm{N}(0, \Gamma)$ is multidimensional Gaussian with scaling matrix $\Gamma = AA^\top$. Then $X \sim \mathrm{S}_k(\alpha, \beta, \Gamma, \delta, \psi)$, $\beta = 0$ is α-stable sub-Gaussian if

$$\begin{aligned} X &\overset{\mathcal{L}}{=} W^{1/2}G + \delta \\ &\overset{\mathcal{L}}{=} W^{1/2}AZ + \delta, \ Z \sim \mathrm{N}(0, I_k) \\ &\overset{\text{def}}{=} AY + \delta, \end{aligned} \tag{3.25}$$

while $Y \sim \mathrm{S}_k(\alpha, 0, I_k, 0)$ is radially symmetric α-stable. The according characteristic function of X is

$$\begin{aligned} \varphi_X(u) &= \int_{-\infty}^{\infty} f_X(x) \exp(iu^\top X) \mathrm{d}x = \mathsf{E}\left(iu^\top X\right) \\ &= \exp\left\{ -\left(\frac{1}{2}u^\top \Gamma u\right)^{\alpha/2} + iu^\top \delta \right\}, \end{aligned} \tag{3.26}$$

$f_X(x)$ as probability density function. $\Gamma \in \mathbb{R}^{k \times k}$ is the positive definite scale matrix and $\delta \in \mathbb{R}^k$ the location vector. The characteristic generator is therefore given by

$$\psi(s, \alpha) = \exp\left\{ -\left(\frac{1}{2}s\right)^{2/\alpha} \right\}. \tag{3.27}$$

This implies that α-stable sub-Gaussian distributions are scale mixtures of multivariate normal distributions (Samorodnitsky and Taqqu, 1994). Note, for $\alpha = 2$, the characteristic function collapses to the Gaussian. For $G \sim \mathrm{N}(0, I_k)$, the characteristic function of Y in Equation 3.25 simplifies to

$$\varphi_Y(u) = \mathsf{E}\left(iu^\top Y\right) = \exp\left(-\gamma^\alpha |u|^\alpha\right). \tag{3.28}$$

For the horizon of the investor, T, the estimated higher sampling frequency log-returns are summed to the chosen frequency:

$$\tilde{X} = TX \ \sim \mathrm{S}_k(\alpha, 0, T\Gamma, T\delta, \psi). \tag{3.29}$$

For subsequent estimation, stability parameter $0 < \alpha \leq 2$, scale Γ and location δ need to be estimated, given that $-1 \leq \beta \leq 1$ can be assumed to be not significantly different from zero.

3.3.4 Parameter estimation

The utilization of α-stable laws implies that fractional moments of random variable X

$$\mathsf{E}|X|^p = \int_{-\infty}^{+\infty} |X|^p f(X) \mathrm{d}X, \tag{3.30}$$

are finite for $0 < p < \alpha$, $p \in \mathbb{R}$ and infinite for $p \geq \alpha$. This implies that for the α-stable Paretian case, representing a slower decay than under the Gaussian, $0 < \alpha < 2$, the second moment $\mathsf{E}|X|^2 = \infty$ and higher moments such as skewness and kurtosis are infinite. For the empirical financial market returns $1 < \alpha < 2$ (see Chapter 3.4), the first moment remains finite. For elliptically α-stable random variable $X \sim \mathrm{S}_k(\alpha, 0, \Gamma, \delta, \psi))$ the expectation is

$$\mathsf{E}\, X = \delta < \infty. \tag{3.31}$$

In general, for univariate α-stable laws the mean is undefined for $\alpha \leq 1$ and $\mathsf{E}\, X = \delta - \beta\gamma \tan\left(\frac{\pi\alpha}{2}\right) < \infty$ for $\alpha > 1$. From the perspective of a data scientist, analyzing the sample, empirical moments are always finite. But under the assumptions of being α-stable distributed, fractional moments with $p \geq \alpha$ have no intrinsic meaning. As shown in Subsection 3.3.1 higher moments behave erratic with increasing data points, contrary to moment convergence under Gaussianity. For portfolio allocation the estimation of location and scale are crucial. Founding on the analysis of Chopra and Ziemba (1993), the mean represents the largest source of error for estimating the portfolio fraction. Their final implication is straightforward: "[...] the bulk of resources should be spent on obtaining the best estimates of expected returns of the asset classes under consideration".

Simulating from the class of elliptically α-stable distributions implies to estimate the stability parameter α, scaling matrix Γ and location δ, given that the skewness parameter β is zero. For the characteristic exponent α the method of Rachev and Mittnik (2000b) is used:

i. Simulate $U_1, \ldots, U_n$ uniformly i.i.d. random variables on the unit hypersphere S^{k-1}.

ii. Estimate the MLE for the index of stability $\hat{\alpha}_i$ (Nolan, 2001) for each i from 1 to n, $U_i^\top X_1, \ldots, U_i^\top X_n$.

iii. Calculate the index of stability by $\hat{\alpha} = n^{-1} \sum_i^n \hat{\alpha}_i$.

By utilizing the MLE for the characteristic exponent α, severe estimation biases from e.g. the Hill estimator (Hill, 1975) are circumvented, see also McCulloch (1997) and Kearns and Pagan (1997). For the proposed semiparametric scaling approximation in Subsection 3.3.5, the estimation of stability α will not be necessary.

Estimating the location vector $\delta \in \mathbb{R}^k$ of multidimensional variable
$X \sim \mathrm{S}_k(\alpha, 0, \Gamma, \delta, \psi)$ is of crucial importance for portfolio allocation, representing *the* driver for asset growth.
From the perspective of information theory, we aim to chose the parameter vector, which maximizes the probability of coming from the empirical data-set. From the perspective of decision theory, this method coincides with the minimization of expected loss under the 0-1 loss function:

$$L(\delta, \hat{\delta}) = 1(\delta \neq \hat{\delta}). \tag{3.32}$$

The according risk function is

$$\mathrm{R}(\delta, \hat{\delta}) = \mathsf{E}\left\{\mathrm{L}(\delta, \hat{\delta})\right\} = \mathsf{E}\left\{1(\delta \neq \hat{\delta})\right\}. \tag{3.33}$$

Consequently, the optimization

$$\delta^* = \underset{\delta^* \in \mathbb{R}^k}{\arg\min} \left[\mathsf{E} \left\{ 1(\delta \neq \hat{\delta}) \right\} \right] \tag{3.34}$$

leads to the common Maximum Likelihood Estimate (MLE). If the loss function is not presumed to be $0-1$ loss, e.g. quadratic, the usual ML estimator may not be suitable. The inadmissability of the sample mean under the Gaussian for dimensions $k > 2$ has been first shown by Stein (1955), leading to the class of shrinkage estimators, starting with James and Stein (1961). An overview of the class of shrinkage estimators is given in Hansen (2015). To our knowledge, those results have not been extended to α-stable laws.

Following Nolan (2013), there are two methods to estimate the scale matrix Γ:

i. Given that X is elliptically α-stable,

$$\forall u,\ u^\top X \sim \mathrm{S}_k \left(\alpha, 0, (u^\top \Gamma u)^{\frac{1}{2}}, u^\top \delta, \psi \right). \tag{3.35}$$

The $k(k+1)/2$ parameters of the scale matrix Γ are estimated by

$$\begin{aligned} \hat{\Gamma}_{j,j} &= \hat{\gamma}_j^2 \\ \hat{\Gamma}_{j,i} &= \tfrac{1}{2} \left\{ \hat{\gamma}^2(1,1) - \hat{\gamma}_i^2 - \hat{\gamma}_j^2 \right\}, \end{aligned} \tag{3.36}$$

where $\hat{\gamma}^2(1,1) = (1,1)^\top (X_j, X_i) = X_j + X_i$ and $\hat{\gamma}_j$ is the univariate scale ML estimate of asset j. Note that $\hat{\Gamma}_{j,i}$ depends solely on directions $(1,1)$,$(1,0)$ and $(0,1)$.

ii. As $\mathsf{E} \left\{ \exp(iu^\top X) \right\} = \exp \left\{ -\gamma(u)^\alpha \right\}$

$$\left\{ -\log \mathsf{E} \exp(iu^\top X) \right\}^{\frac{2}{\alpha}} = u^\top \Gamma u = \sum_i u_i^2 \Gamma_{i,i} + 2 \sum_{i<j} u_i u_j \Gamma_{i,j}, \tag{3.37}$$

so $\Gamma_{i,j}$ can be estimated as linear function via regression, taking more directions into account than the first method.

For the remainder of the paper, the first method is utilized due to its analytical tractability.

3.3.5 Semiparametric scaling approximation

Instead of simulating from the estimated elliptically contoured α-stable distribution, a semiparametric scaling approximation based on higher sampling frequency data is proposed. Simulating from the elliptically α-stable distribution implies that

$$\begin{aligned} &\forall j = 1, \ldots, k \qquad \alpha_j = \alpha,\ 0 < \alpha < 2, \\ &\forall j = 1, \ldots, k \qquad \beta_j = 0. \end{aligned} \tag{3.38}$$

We deal with this drawback by using α-stable properties of the empirical data-set. Assume that the higher sampling frequency data-set $X_t \sim \mathrm{S}_k(\alpha, 0, \Gamma_t, \delta_t, \psi)$ is elliptically α-stable distributed. Then,

i. estimate location δ_t and scale $\Gamma_t = A_t A_t^\top$ of higher frequency returns X_t as proposed in Subsection 3.3.4.

ii. Normalize X_t to radially symmetric $Y \sim \mathrm{S}_k(\alpha, 0, I_k, 0, \psi)$

$$Y = A_t^{-1} X_t - \delta_t. \tag{3.39}$$

iii. Rescale radially symmetric Y to distribution $X \sim \mathrm{S}_k(\alpha, 0, \Gamma, \delta, \psi),\ \Gamma = AA^\top$ with investment horizon T,

$$X = AY + \delta, \tag{3.40}$$

with $\Gamma = T\Gamma_t$ and $\delta = T\delta_t$.

The resulting distribution for horizon T, represented by convoluted higher frequency distributions, is simply an affine transformation of its radially symmetric analogue, given its scaling nature. Given that $\beta = 0$, we can use the potentially different stabilities α_j of the marginals, having no effect on location δ and scale Γ. As the horizon distribution represents a limited number of data points (see Subsection 3.3.1), empirical quantiles $Q_\alpha,\ \alpha < 0.02$ are overestimated, implying that

risk measures for large confidence levels are underestimated. Vice versa, quantiles $Q_\alpha,\ \alpha > 0.98$ are consequently underestimated, see Figure (3.5). By using empirical higher sampling frequency data, we can scale high-frequency events to a manifold of large scale events, which never happened in the original data history of the lower sampling frequency, enriching the tails of the horizon distribution.

3.4 Implementation

3.4.1 Data

The hourly financial stock prices come from Lobster and cover the time span from 2007-06-27 to 2018-05-25, representing 17862 hourly prices per asset. The $k = 14$ assets with a linear payoff structure (stocks) are the stocks with the biggest market capitalization in the NASDAQ 100, representing a technology driven portfolio. A risk-free asset, which can be bought with annual rate $r = 0.01$ is included into the optimization. Relevant asset statistics including Maximum Likelihood Estimates (MLE) under α-stability (Nolan, 2001) are given in Table 3.2.

The assets with a non-linear payoff structure are represented as long put options, written on the stock market index NASDAQ 100. As will be assumed for the representative investor in Subsection 3.4.4, the maturity, and hence the investment horizon T, is chosen to be one year. The prices coming from the ask implied volatilities of the long put options determine the price of the hedge and accordingly the reduction in wealth if the stocks close above the chosen strike levels. For the distribution of wealth in T, the put option price O_T at maturity is given by the inner value

$$O_T = \max\{0; K - S_T\}. \tag{3.41}$$

Solely for evaluating the price of the non-linear assets between $t = 1$ and horizon T, a pricing model is needed.

Table 3.2 Log-return descriptive statistics with Maximum Likelihood Estimates under α-stability, 2007-06-27 to 2018-05-25

	$\mu \times T$	$\sigma \times T^{1/2}$	Skewness	Kurtosis	α	β	$\gamma \times T^{1/\alpha}$	$\delta \times T$
Apple	0.22	0.31	−0.75	41.09	1.33	0.03	0.69	0.08
Adobe	0.16	0.32	−0.97	68.52	1.42	−0.05	0.55	0.33
Amgen	0.11	0.26	1.12	44.33	1.49	0.00	0.39	0.06
Amazon	0.29	0.38	1.79	62.00	1.40	0.03	0.67	0.14
Comcast	0.08	0.30	−0.16	29.31	1.43	−0.02	0.51	0.14
Costco	0.11	0.23	−0.64	32.68	1.44	0.03	0.38	0.04
Cisco	0.04	0.29	−0.94	62.65	1.44	−0.02	0.47	0.09
Gilead	0.11	0.30	−0.94	48.60	1.47	0.02	0.47	−0.01
Intel	0.08	0.28	−0.24	25.55	1.43	−0.01	0.51	0.07
Microsoft	0.11	0.27	−0.05	37.09	1.42	−0.01	0.47	0.09
Nvidia	0.16	0.47	−2.95	134.48	1.40	−0.01	0.86	0.20
Pepsi	0.04	0.18	−0.74	38.56	1.44	0.00	0.30	0.08
Qualcomm	0.03	0.31	−0.31	59.96	1.39	−0.01	0.58	0.07
Texas Instruments	0.10	0.28	−0.71	28.31	1.44	−0.04	0.49	0.25

3.4.2 Stable tests

In order to verify if the class of elliptically α-stable distributions is suitable for the financial assets, the following prerequisites have to be met:

- heavy tails beyond the Gaussian (Leptokurtic behavior),
- linear dependence structure between the margins,
- comparable range of α_j (for simulation),
- skewness parameter β not coherently different from zero.

As examined descriptively in Table 3.2, empirical financial market returns are significantly non-Gaussian. In Figure 3.5, the densities of the normalized log-returns on log-scale are plotted for Gaussian ($\alpha = 2$), Stable ($\alpha = 1.33$), Cauchy ($\alpha = 1$) and the individual assets using Kernel Density Estimates. Within the α-stable framework all examined assets lie between Gaussian and Cauchy, $1 < \alpha < 2$. The α-stable fit for $\alpha = 1.33$ captures the tails adequately, although events are captured, which never took place in the data history. The range of characteristic exponents stands in line with results of Westerfield (1977), McCulloch (1997) or Nolan (2013). The elliptical behavior is assessed by using two dimensional scatter matrices of the empirical log-returns. The significance of the skewness parameters β_j is verified by the utilization of the Fisher information from the MLE. The respective confidence intervals for the individual parameters show that β_j are not consistently different from zero, given a confidence level of 99%. For larger dimensions, Nolan (2013) reaches the same conclusion for the Dow Jones constituents. Making use of the semiparametric scaling approximation implies that there is no need to estimate one specific α for the elliptical α-stable distribution. As $\forall j$ $1 < \alpha_j < 2$ we can deny the null of Gaussianity coherently for the 99% confidence level, speaking in favour of the α-stable hypothesis. As we are interested in the horizon distribution, constituted by the sum of hourly random variables, the generalized CLT is utilized.

3.4.3 Stable estimation

Following Subsection 3.3.4, the parameter estimates for the hourly distribution $X_t \sim \mathrm{S}_k(\alpha, \beta, \Gamma_t, \delta_t, \psi)$, $\beta = 0$ are scaled to the chosen horizon of one year. Exemplary. the semi-log densities for yearly Apple log-returns under Gaussian, Cauchy, Stable and the semiparametric scaling are plotted in Figure 3.6. Additionally, Gaussian scaling, representing the scaling of the hourly distribution utilizing the

square root of time rule under Gaussianity whilst neglecting the CLT, is displayed. In comparison, the semiparametric scaling distribution exhibits heavier tails than under Gaussianity, implying stock market events, which never occurred in the history of the original sampling frequency. The utilized scaling approximation provides the horizon distribution $X_T \sim \mathrm{S}_k(\alpha, 0, \Gamma_T, \delta_T, \psi)$ with location vector $\delta_T = T\delta_t$ and scaling matrix $\Gamma_T = T\Gamma_t$, given that $\beta = 0$.

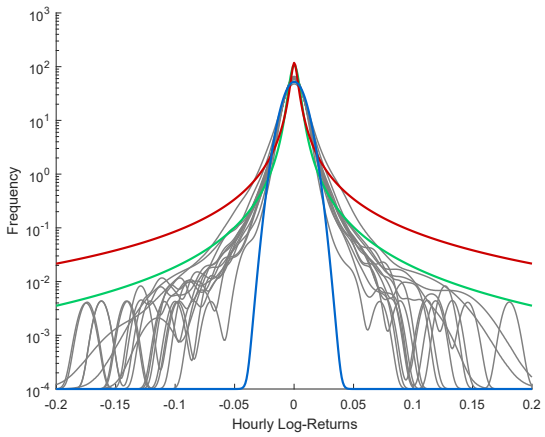

Figure 3.5: Semi-log densities for hourly Apple log-returns for Gaussian, $\alpha = 2$ (blue), Stable, $\alpha = 1.33$ (green), Cauchy, $\alpha = 1$ (red) and financial assets (gray)

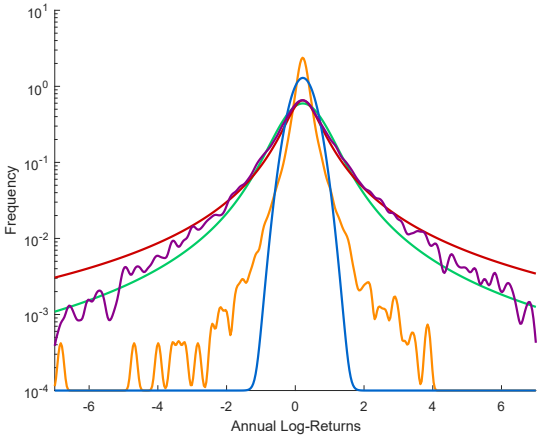

Figure 3.6: Semi-log densities for yearly Apple log-returns for Gaussian, $\alpha = 2$ (blue), Stable, $\alpha = 1.33$ (green), Cauchy, $\alpha = 1$ (red), Gaussian scaling (orange) and semiparametric scaling (violet)

Stable_Kelly_LogDensityFinancials

3.4.4 Portfolio implementation

Exemplary, the representative investor has an investment horizon of one year. According to his client, no more than 20% ($b = 0.20$) of his wealth should be lost given probability $1 - \alpha = 99.5\%$ (Value at Risk). This implies that only $\alpha = 0.5\%$ of the wealth return paths should end below -20%. The investor is able to buy

risk-free bonds with risk free rate $r_f = 1\%$ per year, representing the 15th asset. The maximization problems, without ($k = 15$) and with options ($k = 101$), as a special case of optimization in Equation 3.11, are formulated within the framework of spectral measures. Subject to the VaR$(1-\alpha)$, $\alpha = 0.5\%$ constraint, the Kelly Criterion $G_{\phi_{\text{Elog}}}$ is maximized to achieve the portfolio with the highest growth rate:

$$\begin{aligned} f^* = \underset{f^* \in \mathbb{R}^k}{\arg\max}\ & G_{\phi_{\text{Elog}}}\{W_T(f)\} \\ \text{s.t. } & S_{\phi_{\text{Q}_{0.5\%}}}\left\{1 - \frac{W_T(f)}{W_0}\right\} \leq 0.2, \\ & \sum_{j=1}^{k} f_j \leq 1. \end{aligned} \tag{3.42}$$

Additionally, the client aims to replace the VaR(99.5%) constraint with the Expected Shortfall restriction ES$(1-\alpha)$ in order to account for events beyond the VaR level.

$$\begin{aligned} f^* = \underset{f^* \in \mathbb{R}^k}{\arg\max}\ & G_{\phi_{\text{Elog}}}\{W_T(f)\} \\ \text{s.t. } & S_{\phi_{\text{CTE}_{0.5\%}}}\left\{1 - \frac{W_T(f)}{W_0}\right\} \leq 0.2, \\ & \sum_{j=1}^{k} f_j \leq 1 \end{aligned} \tag{3.43}$$

The resulting discrete wealth return distributions for the VaR restricted portfolios are given in Figure 3.7. Including non-linear instruments into the restricted optimization proves to be beneficial for the Kelly Criterion (Geometric mean), whilst preserving the VaR restriction (see Table 3.3). The protective put strategy allows to reduce probability mass for negative discrete wealth returns. The investment fractions of Table 3.4 show that the decrease in risk free bond for the portfolio with options is equivalent to the option investment.

Replacing the VaR constraint by the ES constraint, indicates that the stock investment is reduced for both cases with and without options, although not substantially. Enriching the ES restricted Kelly portfolios with put options has the

same effect in terms of portfolio fractions as in the VaR restricted case (see Table 3.4) implying a higher geometric mean for the same ES constraint.

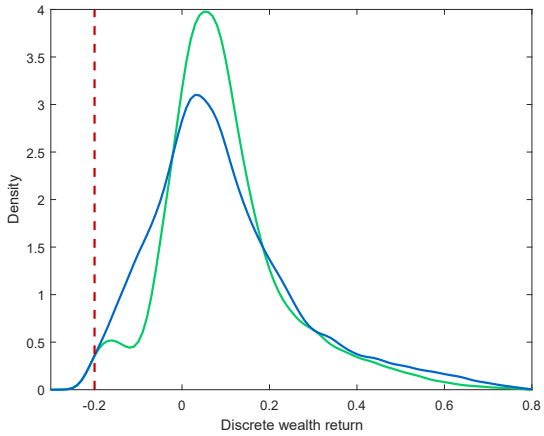

Figure 3.7: Wealth return densities for VaR restricted Kelly optimization without (blue) and with (green) put-options, VaR constraint (red)

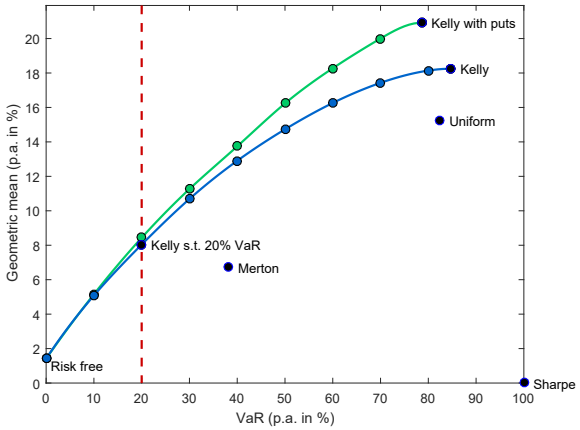

Figure 3.8: Kelly-VaR (99.5%) frontier without (blue) and with (green) options, VaR constraint (red), benchmark portfolios for Uniform, Merton, Sharpe and Kelly (black)

Extending the VaR constraint to the interval $0 \leq b \leq 1$, leads to a series of optimizations for all relevant quantile levels. $b = 0$ represents the risk free portfolio, whereas $b = 1$ implies that the investor can loose all of his fortune, given chosen confidence level.

$$\begin{aligned} f^* = \underset{f^* \in \mathbb{R}^k}{\arg\max} \; & G_{\phi_{\text{Elog}}} \{W_T(f)\} \\ \text{s.t. } & S_{\phi_{\text{Q}_{0.5\%}}} \left\{ 1 - \frac{W_T(f)}{W_0} \right\} \leq b, \; 0 \leq b \leq 1 \\ & \sum_{j=1}^{k} f_j \leq 1 \end{aligned} \tag{3.44}$$

This series of restricted optimizations constitutes the Kelly-VaR frontier (Figure 3.8), in which each point represents a growth-optimal portfolio given quantile

Table 3.3 Discrete wealth return statistics (p.a. in %) for (restricted) Kelly portfolios with and without put options and Sharpe maximizing portfolio

Wealth Statistics	Kelly s.t. VaR $\leq 20\%$	Kelly s.t. VaR $\leq 20\%$ with puts	Kelly s.t. ES $\leq 20\%$	Kelly s.t. ES $\leq 20\%$ with puts	Kelly	Kelly with puts	Sharpe
Geometric mean	8.04	8.44	7.9	8.19	18.25	20.92	-∞
Arithmetic mean	9.64	9.7	9.51	9.45	34.05	32.45	62.08
Standard deviation	17.35	14.45	17.51	14.73	58.42	48.1	124.7
Skewness	1.09	1.06	1.12	1.21	0.65	0.89	1.11
Kurtosis	4.48	5.05	4.53	5.29	3.66	4.77	4.65
Minimum (in %)	−21.86	−21.93	−20.66	−20.79	−95.86	−87.39	−100.00
$\text{CTE}_{0.5\%}$	−20.9	−20.97	−20.00	−20.00	−89.23	−82.41	−100.00
$\text{Q}_{0.5\%}$	−20.00	−20.00	−19.37	−19.29	−84.72	−78.64	−100.00
$\text{Q}_{1\%}$	−19.11	−19.13	−18.55	−18.39	−78.87	−73.43	−100.00
$\text{Q}_{10\%}$	−9.65	−4.67	−9.8	−4.54	−35.1	−9.79	−86.13
$\text{Q}_{50\%}$	6.59	7.39	6.13	6.6	27.17	21.79	40.41
$\text{Q}_{90\%}$	33.25	28.97	33.41	29.03	113.15	97.35	229.96
$\text{Q}_{99\%}$	64.22	55.94	64.76	57.99	198.52	179.02	453.77
Maximum	94.29	87.05	91.78	84.2	330.45	294.65	849.39

(VaR) constraint. The portfolio with (without) options, which can lose at most 20% with 99.5% probability is the portfolio where the green (blue) frontier crosses the quantile constraint (red). Except for the risk-free portfolio, $0 < b \leq 1$, every restricted portfolio with put options outperforms the portfolio without options in terms of the geometric mean. The unrestricted Kelly portfolio exhibits the highest geometric mean possible (18.25%), for a given VaR of 84.72%. Including put options into the unrestricted Kelly optimization increases the geometric mean (20.92%) and reduces the VaR to 78.64% at the same time. Relevant benchmark portfolios such as equally distributed (uniform),

$$f_{\text{Uniform}} = \mathbf{1}\frac{1}{k} \tag{3.45}$$

Table 3.4 Portfolio fractions (in %) for (restricted) Kelly portfolios with and without put options and Sharpe maximizing portfolio

Investment Fractions	Kelly s.t. VaR $\leq 20\%$	Kelly s.t. VaR $\leq 20\%$ with puts	Kelly s.t. ES $\leq 20\%$	Kelly s.t. ES $\leq 20\%$ with puts	Kelly	Kelly with puts	Sharpe
Apple	4.31	4.65	5.24	4.55	19.57	18.76	72.03
Adobe	1.64	3.73	2.16	2.71	10.91	12.15	30.54
Amgen	0.03	1.66	0.65	0.63	9.35	7.05	21.76
Amazon	12.16	10.57	12.43	11.55	24.84	26.30	61.16
Comcast	0.27	0.17	0	0	1.12	0.43	−18
Costco	2.75	1.01	0.27	0.73	11.75	7.34	45.61
Cisco	0	0	0	0	0	0	−53.12
Gilead	0.56	0.39	0.5	0.67	7.97	6.54	14.63
Intel	0.05	0	0	0	0	0	−24.24
Microsoft	0.16	0	0	0	1.57	1.05	5.49
Nvidia	0.31	0.08	0.14	0.54	4.79	5.9	7
Pepsi	0.22	0	0	0	2.96	0	−19.98
Qualcomm	0	0	0	0	0	0	−51.19
Texas Instruments	0.11	0	0.01	0.34	1.05	1.84	8.32
Risk free	77.35	71.96	78.55	73.81	4.06	0	0
Put options	0	5.39	0	4.7	0	12.6	0

the closed-form Merton solution under log-utility and Gaussianity (Merton, 1992),

$$f_{\text{Merton}} = \Sigma^{-1}(\mu - \mathbf{1}r_f) \tag{3.46}$$

and the Sharpe portfolio (Merton, 1972),

$$f_{\text{Sharpe}} = \frac{\Sigma^{-1}(\mu - \mathbf{1}r_f)}{\mathbf{1}^\top \Sigma^{-1}(\mu - \mathbf{1}r_f)} \tag{3.47}$$

are not close to the Kelly-VaR frontier. Specifically the Sharpe investor, who allows for a larger risk constraint b than the unrestricted Kelly solution, should still invest into the growth-optimal portfolio, as the geometric mean of the unrestricted

Kelly portfolio cannot be surpassed. Although the one period Sharpe maximizer obtains a larger arithmetic return, the Kelly (geometric mean) optimization rests on a multi-period investment process. The multi-period investor cannot sustain substantial draw-downs in one period as for the Sharpe portfolio. Given the α-stable process, see Table 3.4, the Sharpe investor goes bankrupt every one-hundred years.

3.5 Conclusion

Whereas the unrestricted Kelly portfolio ensures the asymptotic outperformance of the investor's wealth towards significantly different strategies, the presented model ensures growth-optimal investment subject to personal risk. The constrained optimization is formulated within the framework of spectral measures, inducing quantile (VaR) and Conditional tail expectation (Expected Shortfall) as special cases. In order to allow for an asymmetric wealth distribution, long put options are included into the optimization.

Financial market returns are with large probability non-Gaussian. Founding on the work of Mandelbrot (1963b), it can be observed that the stability parameter α is significantly smaller than two, speaking in favor of the class of α-stable distributions. Given a chosen investment horizon, the distribution of financial market returns is modelled as the sum of hourly random variables. For α-stable laws with $\alpha < 2$, the variance of those random variables is infinite. Hence, the standard CLT does not apply and the generalized CLT of Gnedenko and Kolmogorov (1954b) is applied. For the multidimensional estimation elliptical α-stable distributions, implying a linear dependence structure, are used. Instead of simulating from this class of distributions, a semiparametric scaling approximation is proposed. The resulting annual distribution, represented by convoluted hourly distributions, is simply an affine transformation of its normalized hourly analogue, given its scal-

ing nature.

Heavy tails beyond the Gaussian, linear dependence between the marginals and nonsignificant skewness are empirically supported. Correspondingly, the joint distribution of financial market returns for a specified horizon is estimated by elliptical α-stable distributions utilizing a semiparametric scaling approximation. The portfolio model is implemented for a representative investor with quantile (VaR) constraint. The resulting growth-optimum strategy maximizes the geometric mean, given his risk constraint. Including put options into the optimization levers the portfolio by a suitable protective put strategy, leading to an increased geometric mean for the same risk. For the Kelly-quantile frontier, except for the risk-free portfolio, every restricted portfolio with options outperforms the portfolio without options in terms of the geometric mean.

Chapter 4

Estimating low sampling frequency risk measures by high-frequency data (Univariate)

4.1 Introduction

Value at Risk and Expected Shortfall estimation for high confidence levels, i.e. 99.9% or 99.99% are central parts of risk reporting for banking (Basel III) and insurance (Solvency II) institutions. As the aimed holding periods range from weekly to annual (BIS, 2017), available data-sets for respective sampling frequencies cover a limited amount of data, which is not sufficient in order to model the tails of the return distribution. In this regard, the authors develop a methodology in the α-stable framework, which rescales an empirical high-frequency distribution to a lower sampling frequency. Given more data points provided, the semi-parametric method provides an efficient estimation for risk measures with high confidence levels.

We start to contribute to the literature by formulating a concrete specification of

the sample quantile estimation problem for larger confidence levels given a limited number of observations. Within a simulation study, we show by introducing sample quantile bias and overestimation as a function of the number of data points available, that datasets of ten years history and less do not suffice in order to efficiently estimate quantiles of high confidence (99.9% / 99.99%) for weekly and lower frequencies. The developed frequency rescaling methodology allows to estimate low-frequency risk measures by rescaling high-frequency data, inducing tail events which never occurred in the history of the lower frequency. Furthermore we show that the multifractal scaling law can be mimicked by the frequency rescaling methodology, which employs long-memory GARCH methods. The backtest evaluates the frequency rescaling method in terms of efficiency and coverage and indicates the outperformance over diverse methods for the weekly sampling frequency.

Basel regulations demand that "assumptions made within the internal model are appropriate and do not underestimate risk. This may include the assumption of the normal distribution" (BIS, 2016). As leptokurtic distributions with finite variance - such as Student-t for the daily frequency - are governed by the Central Limit Theorem, the relevant annual distribution would be approximately Gaussian. To circumvent this asymptotic inevitability, returns from different sampling frequencies are analyzed in the framework of α-stable distributions, which have their own domain of attraction and limit theorems (Gnedenko and Kolmogorov, 1954a). Deviating from the Gaussian random walk, first applications in finance are due to Mandelbrot (Mandelbrot, 1963a, 1967) and his student Fama (Fama, 1965), who rest their work on Lévy stable processes (Lévy, 1925). In contrast to the theoretical results of stability under addition (Fama and Roll, 1968), Fama (1976), McFarland et al. (1982), Boothe and Glassman (1987), Dacorogna et al. (2001) as well as Grabchak and Samorodnitsky (2010), show that the stability exponent decreases empirically for higher sampling frequencies. This implies that

high-frequency returns are heavier tailed than low-frequency returns, still differing significantly from Gaussianity. Modeling the dispersion of price increments as a function of its past absolute or squared returns, Bollerslev and Domowitz (1993), Andersen and Bollerslev (1997), Andersen and Bollerslev (1998) and Bollerslev et al. (2000) show that in terms of temporal dependence, higher frequency returns exhibit a longer memory, which call for different function approximations over sampling frequencies. Allowing for hyperbolic decay of the autocorrelation function, long memory can be modelled accordingly (Taylor, 1986). In contrast Mandelbrot and van Ness (1968), Mandelbrot (1982), Mantegna and Stanley (1995) and Xu and Gencay (2003) rest their analysis on the empirical scaling law, which deviates empirically from uni- and mesofractal models. Mandelbrot et al. (1997) and Mandelbrot et al. (1997b) explain the multifractal behavior by modeling the price increments under a subordinated (fractional) Brownian motion. The subordinator is represented by a multifractal measure, which controls tail behavior and long memory (Calvet and Fisher, 2002). On the foundation of frequency dependent stochastic properties in terms of heavy tails, time dependence and multifractality, we analyze ten years of Level 2 tick data for an equally distributed NASDAQ 100 portfolio, which data are gathered from Lobster. For computation, Matlab has been utilized. The paper is organized as follows: The first Chapter formulates the financial model and introduces the sample quantile estimation problem in a simulation study. The high-frequency data-set of NASDAQ portfolio returns is subsequently analyzed over varying sampling frequencies in Chapter two. Given the varying behavior over sampling frequencies, a method to rescale the data-set from high- to low-frequency is introduced in Chapter three. The last chapter verifies the performance of the respective quantile estimates in an in- and out-of-sample backtest.

4.2 Sample quantile distribution

Financial stake-holders are interested in estimating the (conditional) quantile of the wealth distribution for large confidence levels in order to report capital at risk, given a fixed probability of ungovernable events. Although it is possible to estimate risk measures for all confidence levels, given a limited amount of data, we will show that respective risk measures are significantly underestimated. From mathematical statistics it is well-known that the asymptotic distribution of the sample quantile is unbiased and Gaussian by CLT, given stationarity and finite second moments (Ruppert, 2010). Given that empirical time series provide a limited amount of data points, we are going to examine the bootstrapped sample quantile distribution for relevant confidence levels and distribution assumptions.

4.2.1 Financial model

Let $X_t \in \mathbb{R}^k,\ k \in \mathbb{N}^+$ be multidimensional log-returns from distribution P^t, where t indicates the time scale, e.g. days. $\tilde{X}_t$ represents the according discrete returns. For horizon $T \in \mathbb{N}^+$ days, the wealth equation

$$W_T(f_t) = W_0 \prod_{t=1}^{T} \left\{1 + f_t^\top \tilde{X}_t\right\} = W_0 \prod_{t=1}^{T} \left\{f_t^\top \exp(X_t)\right\} \tag{4.1}$$

can be simplified, given constant investment fractions $f \in \mathbb{R}^k$ over time $f_t = f\ \forall t = 0, \ldots, T$.

$$W_T(f) = W_0 \left\{ f^\top \exp\left(\sum_{t=1}^{T} X_t\right)\right\} = W_0 \left\{f^\top \exp(X)\right\},\ X \stackrel{\text{def}}{=} \sum_{t=1}^{T} X_t \tag{4.2}$$

For respective cdf $F_{W_T}(x)$ the spectral measure with weight function $\phi(x)$ is defined through the quantile function $F^{-1}_{W_T}(x) \stackrel{\text{def}}{=} \{x : \mathrm{P}\left(W_T(f_t) \leq x\right) = \tau\},\ \tau \in (0, 1)$.

$$S_\phi\left\{W_T(f_t)\right\} = \int_0^1 \phi(x) F^{-1}_{W_T}(x) \mathrm{d}x \tag{4.3}$$

Within the context of risk measures, the spectral measure will be coherent iff the weight function is positive $\phi(x) \geq 0$, increasing $\phi'(x) \geq 0$ and normalized $\int_0^1 \phi(x) = 1$ (Acerbi, 2002). Two specific risk measures to assess the risk of the portfolio are quantile (VaR) and expected shortfall (ES) constraints:

- S_1 : The quantile (VaR) is a special case of the spectral risk measure from (4.3)

$$\phi_{\mathrm{Q}_\alpha}(x) = \delta(x = \tau), \tag{4.4}$$

  where $\delta(x = \tau)$ is the Dirac delta function, well-known to be a non-coherent risk measure. Further drawbacks of the quantile constraint are treated in Basak and Shapiro (2001).

- S_2 : In contrast, Expected Shortfall is a coherent risk measure representing the average loss beyond a given quantile. Being a special case of the spectral measure, the weight function is given as

$$\phi_{\mathrm{ES}_\tau}(x) = \tau^{-1}\mathbf{1}(x < \tau). \tag{4.5}$$

  Avoiding the weaknesses of VaR, the Basel Committee on Banking supervision proposed to shift the quantitative risk measurement from VaR to expected shortfall (ES) (BIS, 2013, 2016). As ES is shown to be sub-additive and assesses events beyond the quantile, ES is becoming present in the financial industry.

4.2.2 A NASDAQ 100 Investor

Presume a stock-market investor is calculating the weekly VaR of his equally weighted NASDAQ 100 portfolio by using close to ten years of data, representing 549 weekly returns. Accordingly, Gaussian, Student-t ($\nu = 5$) and Stable ($\alpha = 1.7$) distributions are fitted via MLE. The respective quantiles of the discrete wealth

return distribution

$$Q(\tau) = \int_0^1 \delta(x = \tau) F^{-1}(x) dx \tag{4.6}$$

are estimated by the sample quantile

$$\hat{Q}_n(\tau) = \int_0^1 \delta(x = \tau) \hat{F}_n^{-1}(x) dx, \tag{4.7}$$

given in Table 4.1. For the confidence level of 99% the maximum losses of the NASDAQ portfolio returns under standard parametric assumptions are -8.38% (Gaussian), -11.95% (Student-t) and -11.64% (Stable). For confidence intervals larger than 99% the tails of the Stable distribution are heavier than the tails of Student-t and Gaussian. In contrast to the asymptotic Gaussianity of the sample

Confidence $1-\tau$	Gaussian	Student-t	Stable
99%	-6.13	−8.85	−9.12
99.9%	-8.14	−15.14	−29.96
99.99%	-9.76	−23.69	−75.04

Table 4.1: Quantiles (in %) of discrete weekly NASDAQ returns for Gaussian, Student-t ($\nu = 5$) and Stable ($\alpha = 1.7$) for confidence levels 99%, 99.9% and 99.99%

quantile under stationarity and finite second moments we illustrate the effect of limited sample size via bootstrap.

- First, we draw $B = 10^4$ independent bootstrap samples $X^{(1)}, \ldots, X^{(B)}$ of size $n = 569$ from parametric distribution $\hat{F}_{\hat{\theta},n}$,
  $\hat{\theta} = \{$Gaussian, Student-t, $\alpha-$Stable$\}$ and
- accordingly, we calculate $B = 10^4$ quantile estimators $\hat{Q}_{i,n}(\tau) \sim \hat{G}_n$, $i = 1, \ldots, B$ plotted as histograms in Figure 4.1.

We chose the block-length n to be 569 as it represents the number of observations in the weekly frequency of our data-set. The vertical lines in Figure 4.1 represent

the asymptotic quantiles under their parametric assumptions. The bootstrapped quantile distributions for Gaussian, Student-t and Stable are plotted as histograms. Whereas the bootstrapped quantile distributions are unbiased under Gaussianity, the quantile distribution under stable laws is significantly skewed.

In order to evaluate bias and overestimation of the bootstrapped quantile estimates, two measures are introduced as a function of the sample size $n \in \mathbb{N}^+$. The average bias represents the difference between the asymptotic quantile and the bootstrapped quantile:

$$\bar{b}_n = \frac{1}{B} \sum_{i=1}^{B} \left\{ Q(\tau) - \hat{Q}_{i,n}(\tau) \right\}. \tag{4.8}$$

The upper part of Figure 4.2 shows the average bias as function of the sample size. The more leptokurtic the distribution, the more data points are needed in order to obtain an unbiased estimate. Overall, for a confidence level of 99%, 569 observations suffice to provide an unbiased estimate. The quantile overestimation gives the minimum overestimation of the quantile for 5% of the bootstrapped quantiles. In other words, in 5% of the cases the overestimation of the quantile is larger than $O_n(\bar{\tau})$, where

$$O_n(\bar{\tau}) = Q(\tau) - \hat{G}_n^{-1}(\bar{\tau}), \;\; \bar{\tau} = 95\%. \tag{4.9}$$

The lower part of Figure 4.2 shows the quantile overestimation as function of the sample size. The more leptokurtic the distributions, the more data points are needed in order to reduce the overestimation. Overall, for a confidence level of 99% and 569 observations the quantile is overestimated by more than 1% / 1.5% / 2.5% for Gaussian, Student-t and Stable assumption in 5% of the cases. The quantile overestimation implies that the respectively reported VaR estimates are exceeded more often than the confidence level presumes.

For a confidence level of 99.9% 569 observations barely suffice in order to obtain an unbiased estimate (see Figure 4.3). The quantile is overestimated by more than

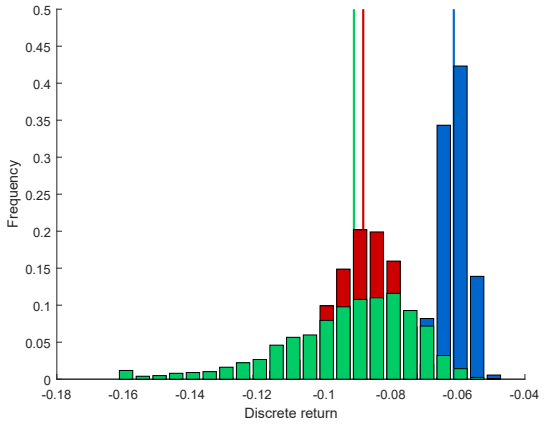

Figure 4.1: Bootstrapped sample quantile histograms for block size $n = 549$ under Gaussianity (blue), Student-t (red) and α-stability (green) with asymptotic quantile as vertical line ($\tau = 1\%$)

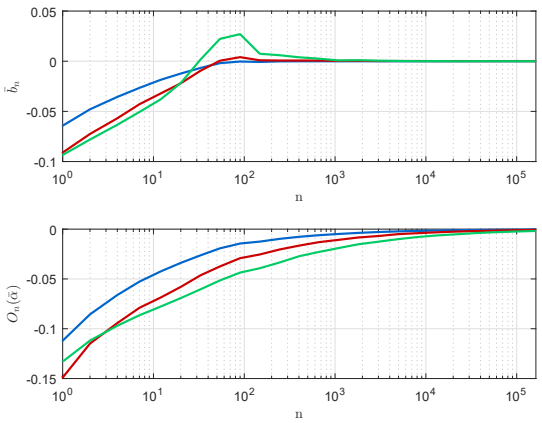

Figure 4.2: (Upper) Sample quantile bias (Lower) $\bar{\tau} = 95\%$ quantile overestimation ($\tau = 1\%$) under Gaussianity (blue), Student-t (red) and α-stability (green)

Quantile_Simulation

1.5% / 5% / 17% for Gaussian, Student-t and Stable assumption in 5% of the cases (see Figure 4.4). In order to obtain a reliable estimate in terms of unbiasedness and overestimation more than 10^4 observations are necessary. For a confidence level of 99.99% 569 observations do not suffice in order to obtain an unbiased estimate. In that respect, the quantile is overestimated by more than 3% / 13% / 62% for Gaussian, Student-t and Stable assumption in 5% of the cases. To obtain a reliable estimate in terms of unbiasedness and overestimation more than 10^5 observations are necessary.

The implication is that although risk measures can always be calculated for high confidence levels, the limited amount of data points leads inevitably to underestimated risk estimates as shown for the special case of the quantile (VaR). For Expected Shortfall the degree of unbiasedness and overestimation is even aggravated due to the conditional formulation of the quantile. Straight-forward, financial in-

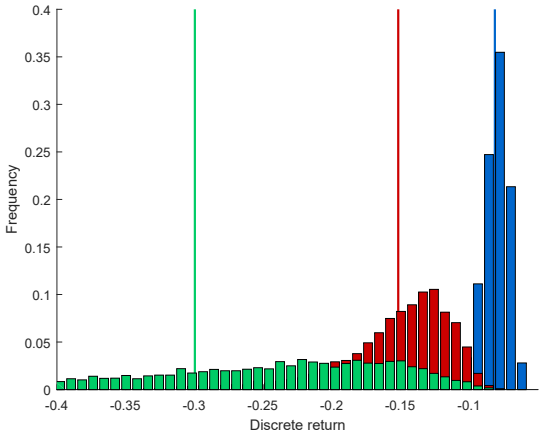

Figure 4.3: Bootstrapped sample quantile histograms for block size $n = 549$ under Gaussianity (blue), Student-t (red) and α-stability (green) with asymptotic quantile as vertical line ($\tau = 0.1\%$)

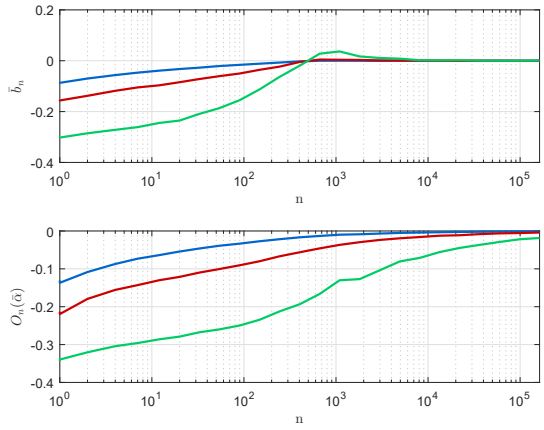

Figure 4.4: (Upper) Sample quantile bias (Lower) $\bar{\tau} = 95\%$ quantile overestimation ($\tau = 0.1\%$) under Gaussianity (blue), Student-t (red) and α-stability (green)

Quantile_Simulation

stitutions should only report risk measures for those confidence levels, which can be estimated efficiently for available data. Within this paper, we are going to describe how to filter and rescale data from higher frequencies, implying significantly more data points, in order to estimate lower frequency risk measures.

4.3 Data

4.3.1 High-frequency data

The data-set represents transaction level data (Level 1) from 2007-06-27 till 2018-05-25 for the largest thirteen stocks by market capitalization gathered from Lobster. By utilizing the previous-tick method, each day gives 390 trading minutes, representing 6.5 hours of trading from 09:30 a.m. till 04:00 p.m., see also Da-

corogna et al. (2001). After transforming the price data to log-returns, the returns are aggregated to their respective frequencies, up to one week, representing 549 weeks (see Table 4.2). Two excerpts of the data-set are given in Figure 4.5. The blue line represents minute, the red line hourly, the orange line daily and the violet line weekly prices. Accordingly, we create an equally weighted portfolio of those NASDAQ 100 stocks to represent a market value driven portfolio.

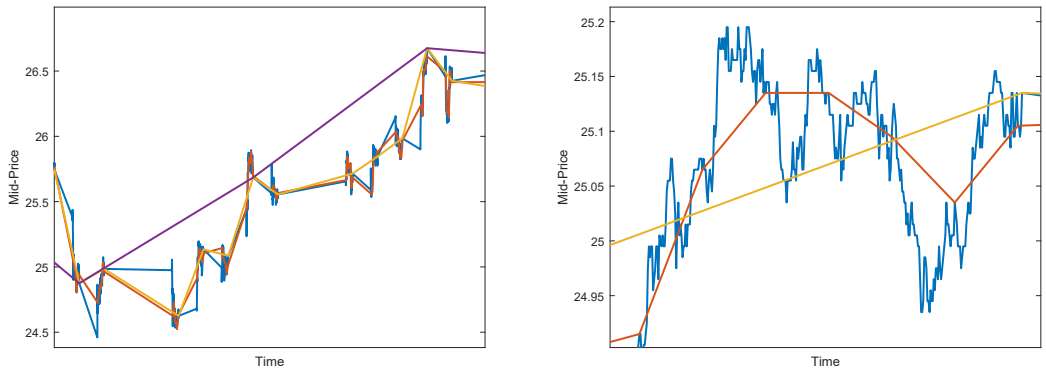

Figure 4.5: Excerpt of NASAQ portfolio mid-price data for minute (blue), hourly (red), daily (orange) and weekly (violet) frequency

Frequency	1 min	1 hour	1 day	1 week
Data points	1071720	17862	2748	549

Table 4.2: Number of closing prices for different sampling frequencies

4.3.2 Stock return characteristics

Data series from various fields of research, i.e. finance, economics, biology and physics, share the same characteristics over time and frequency. Starting with Mandelbrot (1963a), log-transformations of cotton prices and later wheat prices, railroad stocks and financial rates (Mandelbrot, 1967) deviated from the Gaussian

by exhibiting heavier tails than presumed. An overview over various heavy-tailed models in finance is given in Rachev (2003). Stable laws specifically are treated in Zolotarev (1986), Samorodnitsky and Taqqu (1994) and Nolan (2017). For financial returns, the degree of leptokurticity decreases with increasing sampling frequency (Fama, 1976; McFarland et al., 1982; Boothe and Glassman, 1987; Dacorogna et al., 2001; Grabchak and Samorodnitsky, 2010). For NASDAQ portfolio log-returns, the sample kurtosis decreases from 519.64 for 1-minute returns to 9.45 for weekly returns (see Figure 4.6). Hence, the leptokurtic behavior of portfolio log-returns decrease over the sampling frequencies, still differing significantly from the Gaussian assumption.

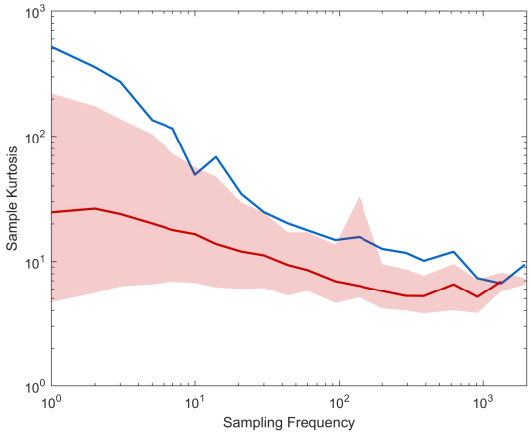

Figure 4.6: Sample kurtosis over frequency (blue) with bootstrapped kurtosis distribution (red) for 549 data points (weekly frequency)

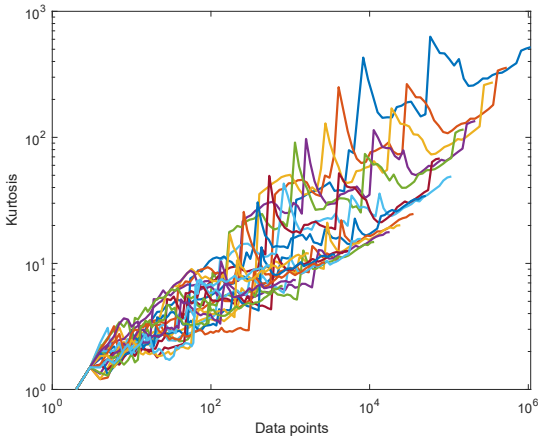

Figure 4.7: Sample kurtosis for increasing data points for frequencies from 1 minute to 1 week

In order to verify if the sample kurtosis of stock returns for lower frequencies is underestimated due to the lack of data points, we sample 10^4 blocks of 549 minutes, hours and days and plot the resulting sample kurtosis confidence interval (confidence level 95%) in Figure 4.6. If 549 data points would suffice on average,

then the mean of the bootstrap distribution should coincide with the empirical kurtosis over the sampling frequency. But the empirical sample kurtosis increases exponentially with increasing frequencies, whereas the average bootstrapped kurtosis remains in a range of six to eleven. For frequencies higher than 10-minutes, even the upper 95% confidence bound of the bootstrapped kurtosis lays below the empirically observed kurtosis. The implication is that 549 data points do not suffice to replicate the sample kurtosis of higher-frequency returns, inducing evidence that the kurtosis of the weekly frequency is underestimated. Indeed, the sample kurtosis increases with increasing data points. In Figure 4.7 the sample kurtosis is plotted as function of number of data points, used for the respective frequencies. The more data points are used, the higher is the sample kurtosis, arguing against the finiteness of sample kurtosis. Accordingly, we resort to analyze the varying distributions over frequency in the α-stable framework for section 4.4, as α-stable distributions belong to their own domain of attraction, allowing for infinite moments such as variance, skewness or kurtosis.

Volatility clustering is the empirical observation that "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes" (Mandelbrot, 1963a). Subsequently, two branches in the analysis of time-dependence developed: On the one hand the (FI)GARCH approach of Engle (1982), Bollerslev (1986) and Baillie et al. (1996), who aim to model the variance as a linear function of the past squared daily returns and on the other hand, the multifractal approach of Mandelbrot et al. (1997), Fisher et al. (1997), Calvet et al. (1997) and Calvet and Fisher (2002). For higher sampling frequencies than daily, see for example Bollerslev and Domowitz (1993), Andersen and Bollerslev (1997), Andersen and Bollerslev (1998) and Bollerslev et al. (2000). For increasing frequency, temporal dependence in absolute and squared returns increases (see Figure 4.8). Specifically, 5-minute and 1-hour absolute NASDAQ portfolio returns exhibit strong intraday seasonalities, which are not captured by standard time-

series models.

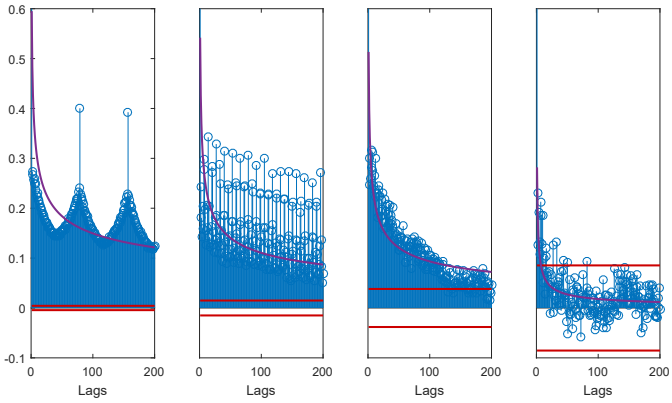

Figure 4.8: Sample autocorrelation function of absolute returns for 5-minute, hourly, daily and weekly frequencies with confidence bounds (red) and hyperbolic fit (violet)

The empirical scaling law states that the mean absolute (squared) returns, as functions of their time intervals, are proportional to a power of the interval size (Mandelbrot, 1982; Mantegna and Stanley, 1995; Mandelbrot et al., 1997b; Calvet and Fisher, 2002; Xu and Gencay, 2003). Starting from the self-affine process $\{X_t\}$, $t \geq 0$ with Hurst exponent $H > 0$ and $c > 0$,

$$X_{ct} \stackrel{\mathcal{L}}{=} c^H X_t \tag{4.10}$$

$$\mathsf{E}\{|X_{ct}|^p\} = c^{Hp}\,\mathsf{E}\{|X_t|^p\} \tag{4.11}$$

the scaling relation-ship in moments of order $p \in \mathbb{R}$ is derived. For $c(p) = \mathsf{E}\{|X_t|^p\}$ and $D(p) = H(p)p$ Mandelbrot et al. (1997) define a fractal process in terms of its moments, remaining graphically tractable (see Figure 4.9).

$$\mathsf{E}(|X_t|^p) = c(p)t^{D(p)} \tag{4.12}$$

For normalization in p, raise the scaling law of Equation 4.12 to the power of $1/p$,

giving

$$\mathsf{E}\left(|X_t|^p\right)^{1/p} = \left\{c(p)t^{D(p)}\right\}^{1/p} \tag{4.13}$$

$$\Rightarrow \frac{1}{p}\log \mathsf{E}\left(|X_t|^p\right) = \frac{1}{p}\log c(p) + H(p)\log t. \tag{4.14}$$

If the absolute moments would scale with a unique Hurst exponent $0.5 \leq H \leq 1$ for all powers p, the underlying process would come from a Fractional Browian Motion (Unifractal). For the Lévy stable motion the stability exponent would imply $H = 1/\alpha$ for $p \leq \alpha$ and $H = 1/q$ for $p > \alpha,\ 0 < \alpha \leq 2$ (Mesofractal). But Figure 4.9 and 4.10 indicate that the Hölder exponents vary with increasing order of the moment p (Multifractal). Müller et al. (1990) argue that the empirically observed scaling law can only be explained by varying distributions for different time intervals, leading to subordinated (Fractional) Brownian Motions, see Mandelbrot et al. (1997b). The Hölder exponents are estimated by (log-log) linear regression, see Equation (4.14) and the generalized Hurst exponent by Matteo et al. (2005).

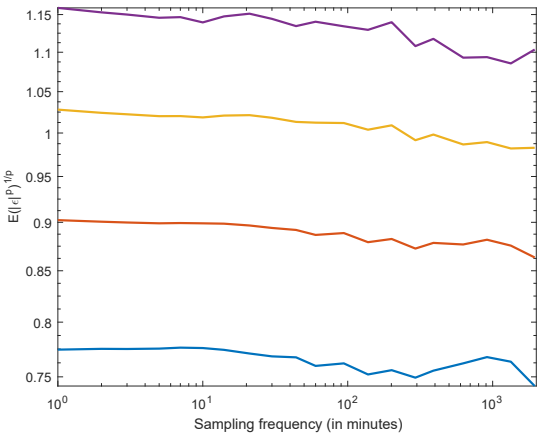

Figure 4.9: Empirical scaling law for NASDAQ portfolio returns formulated as powers $p = \{0.5, 1, 1.5, 2\}$ of absolute moments (from the bottom up)

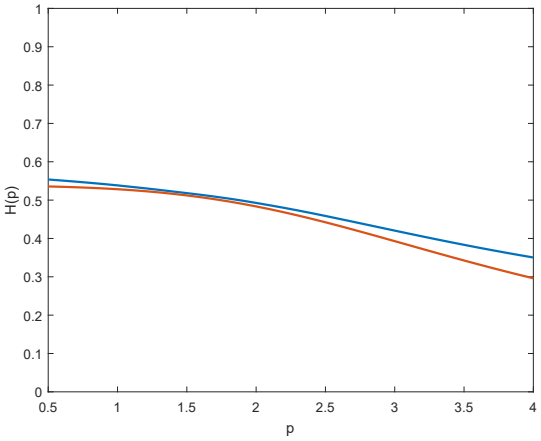

Figure 4.10: Hölder exponents $H(p)$ of NASDAQ portfolio returns for absolute moments of order p, estimated by log-log regression (blue) and generalized Hurst exponent (red)

4.4 Estimation

4.4.1 Mesofractality

A one-dimensional random variable $X \sim \mathrm{S}(\alpha, \beta, \gamma, \delta)$ will be α-stable distributed with parameters $0 < \alpha \leq 2,\ -1 \leq \beta \leq 1,\ \gamma \geq 0$ and $\delta \in \mathbb{R}$ (Nolan, 2017; Cizek et al., 2011), if

$$X \stackrel{\mathcal{L}}{=} \left\{ \begin{array}{r} \gamma Z + \delta,\ \alpha \neq 1 \\ \gamma Z + (\delta + \beta \frac{2}{\pi} \gamma \log \gamma),\ \alpha = 1 \end{array} \right\}, \tag{4.15}$$

where $\mathrm{S}(Z \mid \alpha, \beta, 1, 0)$ represents the standard stable form. As only special cases of stable distributions are available as real-valued densities (Gaussian, Cauchy and Lévy), α-stable distributions are expressed as Fourier transforms of the characteristic function $\varphi_X(u)$.

$$\mathrm{S}(X \mid \alpha, \beta, \gamma, \delta) = \frac{1}{2\pi} \int \varphi_X(u) - \exp(-iuX) du \tag{4.16}$$

The according characteristic function representation is given by

$$\log \varphi_X(u) = \left\{ \begin{array}{l} iu\delta - \gamma^\alpha |u|^\alpha \left\{ 1 + i\beta \tan\left(\frac{\alpha\pi}{2}\right) (\operatorname{sign} u) \right\},\ \alpha \neq 1 \\ iu\delta - \gamma |u| \left\{ 1 + i\beta \frac{2}{\pi} (\operatorname{sign} u) \log(|u|) \right\},\ \alpha = 1. \end{array} \right. \tag{4.17}$$

Scale invariance under addition implies that for the sum of α-stable variables $X_t \sim \mathrm{S}(\alpha, \beta, \gamma, \delta)$, $t = 1, \ldots, T$

$$X_1 + X_2 + \ldots + X_T = \sum_{t=1}^{T} X_t = X \sim \mathrm{S}(\alpha, \beta, T^{\frac{1}{\alpha}} \gamma, T\delta). \tag{4.18}$$

The characteristic function of X is consequently given by

$$T \log \varphi_X(u) = \left\{ \begin{array}{l} iu(T\delta) - T\left(\gamma^\alpha\right) |u|^\alpha \left\{ 1 + i\beta \tan\left(\frac{\alpha\pi}{2}\right) (\operatorname{sign} u) \right\},\ \alpha \neq 1 \\ iu(T\delta) - T\gamma |u| \left\{ 1 + i\beta \frac{2}{\pi} (\operatorname{sign} u) \log(|u|) \right\},\ \alpha = 1. \end{array} \right. \tag{4.19}$$

Gnedenko and Kolmogorov (1954a) prove that the limiting distribution of T i.i.d. α-stable random variables, $0 < \alpha \leq 2$ is

$$a_T \left(\sum_{t=1}^{T} X_t \right) - b_T \stackrel{\mathcal{L}}{\to} \mathrm{S}(\alpha, \beta, 1, 0), \tag{4.20}$$

where $a_T > 0$ and $b_T \in \mathbb{R}$. The special case of the Generalized Central Limit Theorem (GCLT) is the standard CLT for $\alpha = 2$, $\beta = 0$, $\gamma = \frac{\sigma}{\sqrt{2}}$ and $\delta = \mu$, given $a_T = \frac{1}{\sigma\sqrt{T}}$ and $b_T = \frac{\sqrt{T}\mu}{\sigma}$. In general, for $0 < \alpha \leq 2$,

$$T^{-\frac{1}{\alpha}} \sum_{t=1}^{T} (X_t - \delta) \xrightarrow{\mathcal{L}} \mathrm{S}(\alpha, 0, \gamma, 0). \tag{4.21}$$

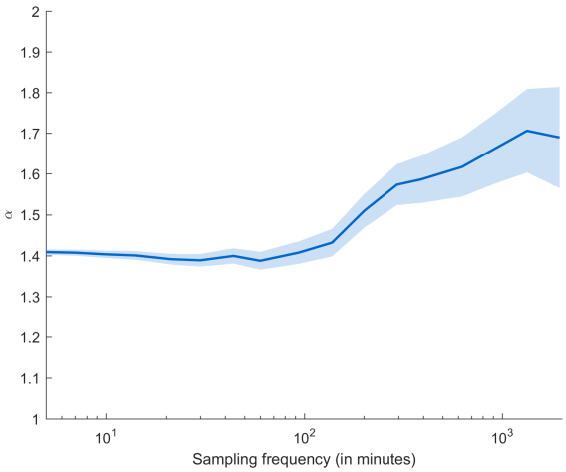

Figure 4.11: Stability index α over sampling frequencies with 95% confidence bounds by MLE

If the scaling exponent α would be constant over the sampling frequency, the sum of the higher frequency returns under α-stability, see Equation 4.1, could be modelled under one specific stable distribution (Fama and Roll, 1968). Figure 4.11 shows the MLE of the stability index α with respect to the sample frequency, including the 95% confidence intervals from the numerical Fisher information (Nolan, 2001). For all sampling frequencies, the respective distributions are more leptokurtic than under Gaussianity ($\alpha = 2$) and more platykurtic than under the Cauchy assumption ($\alpha = 1$). In contrast to the analysis of higher moments, i.e. kurtosis, this class of Stable Paretian distributions $1 \leq \alpha \leq 2$ provides a well-defined framework in order to assess the tails of the distributions of different sampling frequencies. But,

as argued in the context of scaling laws of Subsection 4.3.2, varying distributions over sampling sequences are observed. Consequently, the mesofractal assumption of NASDAQ portfolio returns has to be denied. If the distribution of the higher frequency returns X_t would be modelled under finite variance, such as generalized hyperbolic distributions with normal-inverse Gaussian (NIG) (Hartmann et al., 2010) or Student-t (Chen et al., 2010) as special cases, the horizon distribution, which is heavy tailed by empirical observation (see Figure 4.6), would be asymptotically Gaussian by the standard Central Limit Theorem.

4.4.2 Filter for seasonality and time-dependence

In order to examine the intraday seasonalities of Figure 4.8, we plot the absolute 1-minute returns over the course of the day in Figure 4.12. As the apparent convex shape, see also Engle and Sokalska (2012), is not covered by economic theory, the literature proposed to estimate these intraday seasonalities by universal function approximators. Whereas Giot (2005) models the intraday patterns with cubic splines, Andersen and Bollerslev (1997, 1998) use flexible fourier forms. We follow Andersen et al. (2003) and Engle and Sokalska (2012) by averaging the absolute returns over each minute $k = 1, \ldots, 390$ of the day.

$$s_k = \frac{1}{T} \sum_{t=1}^{T} |X_{t,k}|, \;\; k = 1, \ldots, 390 \tag{4.22}$$

After normalizing the raw one-minute returns by

$$Z_{t,k} = \frac{X_{t,k}}{\sqrt{s_k}}, \;\; k = 1, \ldots, 390 \tag{4.23}$$

the according lower frequencies are calculated and the sample autocorrelation functions of absolute returns are plotted in Figure 4.13. The ACFs of the deseasonalized absolute returns in Figure 4.14 indicates the hyperbolic decay which can be observed for sampling frequencies up to daily, see also Taylor (1986), Robinson

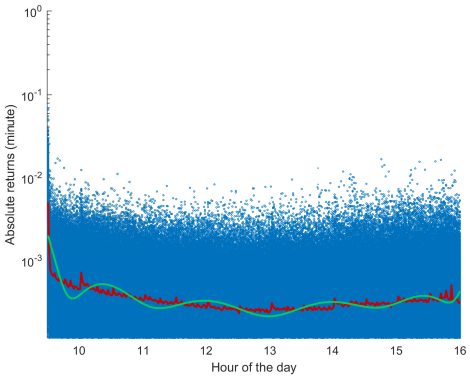

Figure 4.12: Absolute returns (minute) with average (red) over hour of the day and polynomial fit (green)

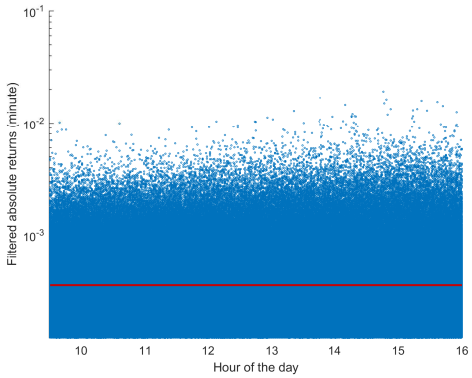

Figure 4.13: Filtered absolute returns (minute) with average (red) over hour of the day

(1991), Ding et al. (1993) and Andersen and Bollerslev (1997). Accordingly, we model the observed phenomena of heavy tails, seasonality and long-range dependence by employing the FIGARCH methodology (Baillie et al., 1996) for seasonally filtered returns (Andersen and Bollerslev, 1998; Müller et al., 1990) under stable laws (Paolella et al., 2002) in order to explain the instability in distributions, as observed in Subsection 4.3.2.

The Stable paretian power GARCH process $\mathrm{S}^h_{\alpha,\beta}\,\mathrm{GARCH}(r,s)$ with $r,s \in \mathbb{N}^+,\ h \in \mathbb{R}$ and seasonality component s_k is given by

$$X_t = \delta + s_k \gamma_t \varepsilon_t,\ \ \varepsilon_t \sim \mathrm{S}_{\alpha,\beta}(1,0) \tag{4.24}$$

$$\gamma_t^h = \theta_0 + \sum_{i=1}^{r} \theta_i |\varepsilon_{t-i}|^h + \sum_{j=1}^{s} \phi_j \gamma_{t-j}^h \tag{4.25}$$

$$= \theta_0 + \theta(L)\varepsilon_t^h + \phi(L)\gamma_t^h. \tag{4.26}$$

The according GARCH equation can be rewritten in lag polynomial form:

$$\{1 - \theta(L) - \phi(L)\}\,\varepsilon_t^h = \theta_0 + \{1 - \phi(L)\}\left\{\varepsilon_t^h - \gamma_t^h\right\}. \tag{4.27}$$

In order to allow for a slower decay than exponential, the fractional difference

operator $(1-L)^d$, $0 < d < 1$ is introduced, obtaining the fractionally integrated GARCH (FIGARCH) equation:

$$\{1-\theta(L)-\phi(L)\}(1-L)^d\varepsilon_t^h = \theta_0 + \{1-\phi(L)\}\left\{\varepsilon_t^h - \gamma_t^h\right\}. \tag{4.28}$$

In contrast to the special cases of GARCH ($d=0$) and IGARCH ($d=1$), Davidson (2004) shows that this class of processes is able to reproduce more flexible temporal dependencies, i.e. long memory. As GARCH processes are modelled separately for each sampling frequency Mandelbrot et al. (1997) argues that this family of fractionally integrated models is neither self-affine nor scale consistent. Still, Fisher et al. (1997) find evidence that the class of FIGARCH-models can mimic multifractality.

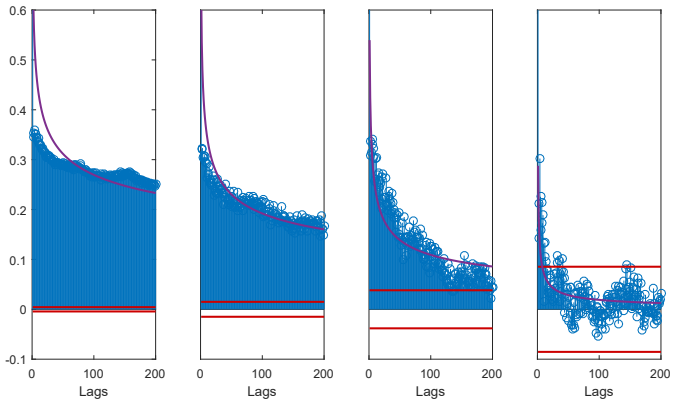

Figure 4.14: Sample autocorrelation function of deseasonalized absolute returns for 5-minute, hourly, daily and weekly frequencies with confidence bounds (red) and hyperbolic fit (violet)

4.4.3 Frequency rescaling

Figure 4.15 shows the ML estimates of the stability parameter α after deseasonalizing with (green) and without (blue) FIGARCH(1,1) filter. Comparable to Figure 4.11, deseasonalizing price increments without FIGARCH(1,1) filter gives again

evidence for different generating distributions for different sampling frequencies. After accounting for temporal dependence via FIGARCH(1,1) filter, the stability parameter remains to be constant for sampling frequencies larger than five minutes. The specific parameters of the FIGARCH(1,1) models for the different sampling frequencies are available on request. Here, the increase of the stability index with decreasing sampling frequency can be explained to a large extent by intraday seasonality and time dependence. For higher frequencies than five minutes, microstructure effects lead to a overestimated deviations and hence a smaller stability index (Zumbach et al., 2002; Chaboud et al., 2010).

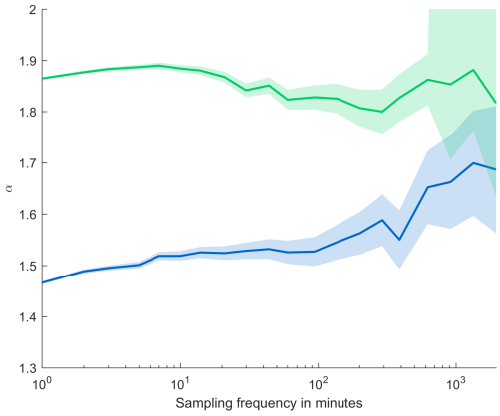

Figure 4.15: Stability parameter of innovations ε_t over frequency with (blue) and without (green) FIGARCH filter including 95% confidence bands from the MLE

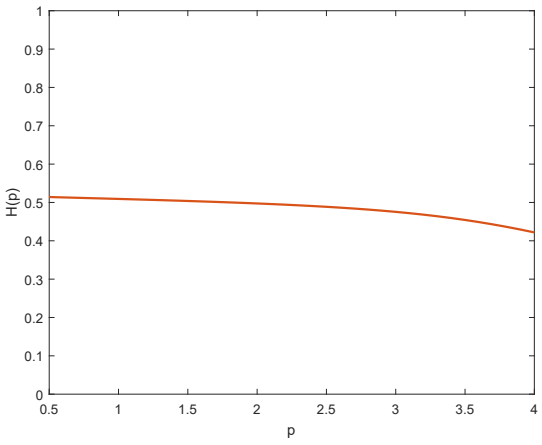

Figure 4.16: Hölder exponents $H(p)$ of FIGARCH-residuals for absolute moments of order p, estimated by log-log regression (blue) and generalized Hurst exponent (red)

The result is supported from the perspective of scaling laws, see Equation 4.14. The observation of multifractality is not evident in the residuals of the FIGARCH(1,1) model. Figure 4.16 plots Hölder exponents, which remain to be constant over the residuals moments of order p. We have shown that the filtered log-returns give

evidence for mesofractality,

$$\varepsilon_t \sim \mathrm{S}_{\gamma,\delta}(\alpha, \beta) \tag{4.29}$$

for all sampling frequencies larger than five minutes, which makes it possible to rescale between different frequencies under α-stability, see Equation 4.18. The proposed semiparametric rescaling method rests on the higher frequency data-set itself, but uses the α-stable assumption beneath:

i. Filter higher frequency returns X for intraday seasonality and time dependence.

   - Let higher frequency returns $X \sim \mathrm{S}(\alpha_X, \beta_X, \gamma_X, \delta_X)$,
   - higher frequency residuals $\varepsilon_X \sim \mathrm{S}(\alpha_\varepsilon, \beta_\varepsilon, \gamma_\varepsilon, \delta_\varepsilon)$ and
   - let lower frequency returns $Y \sim \mathrm{S}(\alpha_Y, \beta_Y, \gamma_Y, \delta_Y)$.

ii. Evaluate if $\alpha_\varepsilon = \alpha_Y \overset{\text{def}}{=} \alpha$, $\beta_\varepsilon = \beta_Y \overset{\text{def}}{=} \beta$.

iii. Normalize higher frequency residuals (for auxiliary models).

$$Z = \frac{\varepsilon_X - \delta_\varepsilon}{\gamma_\varepsilon}, \; Z \sim \mathrm{S}(\alpha, \beta, 1, 0) \tag{4.30}$$

iv. Rescale normalized residuals for lower frequency with drift $\delta_{Y_{\mathrm{FRM}}} = T\delta_X$ and scale $\gamma_{Y_{\mathrm{FRM}}} = T^{\frac{1}{\alpha_X}}\gamma_X$.

$$Y_{\mathrm{FRM}} = \delta_{Y_{\mathrm{FRM}}} + \gamma_{Y_{\mathrm{FRM}}} Z, \; Y_{\mathrm{FRM}} \sim \mathrm{S}(\alpha, \beta, \gamma_{Y_{\mathrm{FRM}}}, \delta_{Y_{\mathrm{FRM}}}) \tag{4.31}$$

v. Estimate the risk measure from the nonparametric, rescaled distribution $F_{Y_{\mathrm{FRM}}}$, see Equation 4.3.

$$\mathrm{S}_\phi(Y_{\mathrm{FRM}}) = \int_0^1 \phi(x) F^{-1}_{Y_{\mathrm{FRM}}}(x) dx \tag{4.32}$$

By obtaining a lower-frequency distribution from high-frequency data, the problem of insufficient data points from high-confidence risk measures of Section 4.2.2 is addressed. For the aimed weekly frequency, we rescale 5-minute returns to the weekly frequency. For aimed frequencies higher than weekly, the time-dependence structure of the lower frequency would have to be included. Figure 4.17 compares the empirical lower frequency distribution with distribution F

$$Y_{\text{EMP}} \sim \text{F} \tag{4.33}$$

with Gaussian scaling given empirical higher frequency returns $X \sim \text{N}(\mu_X, \sigma_X)$,

$$Y_{\text{Gauss}} = T\mu_X + T^{\frac{1}{2}}\sigma_X Z, \tag{4.34}$$

$$Y_{\text{Gauss}} \sim \text{N}(T\mu_X, T^{\frac{1}{2}}\sigma_X), \;\; Z = \frac{X - \mu_X}{\sigma_X} \sim \text{N}(0,1). \tag{4.35}$$

Stable scaling given empirical higher frequency returns $X \sim \text{S}(\alpha_X, \beta_X, \gamma_X, \delta_X)$,

$$Y_{\text{Stable}} = T\delta_X + T^{\frac{1}{\alpha_X}}\gamma_X Z, \tag{4.36}$$

$$Y_{\text{Stable}} \sim \text{S}(\alpha_X, \beta_X, T\delta_X, T^{\frac{1}{\alpha_X}}\gamma_X), \;\; Z = \frac{X - \delta_X}{\gamma_X} \sim \text{S}(\alpha_X, \beta_X, 0, 1) \tag{4.37}$$

and the frequency-rescaled distributions (FRM). Whereas the empirical distribution is constituted by 549 data points, the scaled (Gaussian, Stable and FRM) distributions utilize $214,343$ rescaled 5-minute returns respectively. The distribution under Gaussian scaling, driven by the standard CLT, is lighter tailed ($\alpha = 2$) than the empirical distribution. Under stable scaling ($\alpha = 1.47$), driven by the generalized CLT, the weekly distribution is heavier tailed than the empirical distribution. The FRM provides a distribution, which represents a tradeoff between being Gaussian and Stable scaling, still being significantly heavier tailed than the empirical distribution ($\alpha = 1.86$). Albeit coming from the same data-set, the scaling methods allow for sampling positive and negative events, which never occurred in the original weekly data history. In the backtest we aim to show that the Gaussian scaling is too light tailed to approximate the tails properly, whereas the Stable scaling ought to be too heavy tailed.

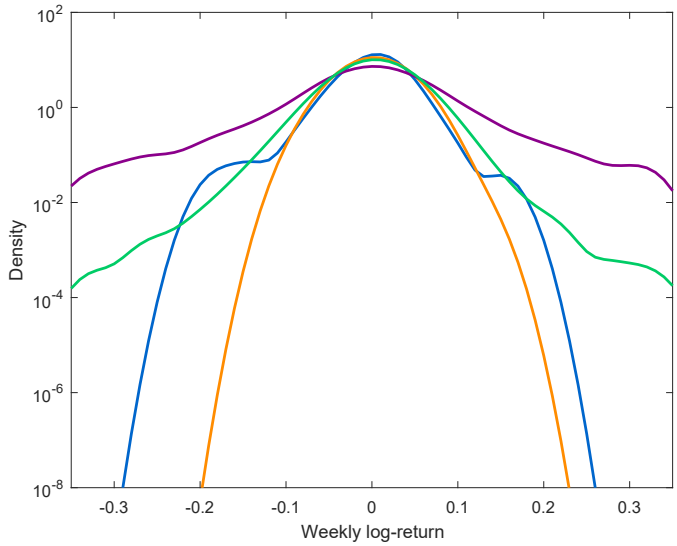

Figure 4.17: Kernel Density Estimates for weekly return distributions: Empirical (blue), Gaussian scaling (orange), Stable scaling (violet) and FRM (green)

4.5 Implementation: VaR

4.5.1 Backtest methodology

Due to transaction costs investors and other stakeholders are not rebalancing their portfolios at high-frequencies. Consequently institutional investors and also regulators are naturally interested in low-frequency risk measures at weekly, quarterly or annual time scale. Here we focus on the 5-day "minimum liquidity horizon" to guarantee a sufficient amount of data, e.g. for the backtest (BIS, 2017). With the spectral risk measure 4.3 the exceedance process $I_t,\ t = 1, \ldots, T$ is defined as

$$I_t = \begin{cases} 0, & X_t > S_{\phi_\alpha}() \\ 1, & X_t \leq S_{\phi_\alpha}(). \end{cases} \tag{4.38}$$

Given i.i.d. data X_t, the unconditional coverage is given by

$$\mathsf{E}\,(I_t) = \tau. \tag{4.39}$$

Statistical coverage tests by Kupiec (1995) and Christoffersen (1998) utilize that under such an i.i.d. assumption $I_t \sim \text{Bern}(\tau)$ and accordingly $\sum_{t=1}^{T} I_t \sim \text{Bin}(T + 1, 1 - \tau)$. A test statistics to evaluate

$$H_0 : \hat{\tau} = \tau \text{ vs.} \tag{4.40}$$

$$H_1 : \hat{\tau} \neq \tau. \tag{4.41}$$

is easily constructed. Alternatively, BIS (1996) propose to use a traffic light approach, which we will follow here. We are going to compare the proposed frequency rescaling method (FRM) of Section 4.4.3 with the VaR forecasts of the following models for the weekly sampling frequency: Under the independence assumption Gaussian, Student-t, Stable and nonparametric VaR estimates are directly calculated from the according ML fits given weekly data. Accounting for time-dependence in the data (see Figure 4.8), GARCH(1,1) and FIGARCH(1,1) risk estimates are obtained for weekly and 10-minute data. Additionally, the VaR forecasts for the realized GARCH(1,1) and realized FIGARCH(1,1) are included. As argued in Section 4.2, 549 observations from the weekly frequency do not suffice to construct reliable VaR estimates, especially for large confidence levels. This holds for the backtest of the weekly VaR. Given $549 - h,\ h \in \mathbb{N}^+$ testing weeks, where $h = 41$ is the number of periods used to initially estimate the VaR levels, the exceedance probability cannot be sufficiently estimated for large confidence levels. But as empirical data remain the only viable foundation for the backtest, the in-sample and out-of-sample exceedances (in%) are given in Table 4.3 and 4.4 for confidence levels 95%, 99% and 99.9%. The unconditional coverage holds for all respective confidence levels, in- and out-of-sample, for the 3) Stable, 6) Weekly FIGARCH and 11) Frequency Rescaling models.

Among the models satisfying the theoretically presumed confidence levels, a smaller level of VaR is beneficial for banks, insurances and other financial institutions. Accordingly Figure 4.18 and 4.19 show in- and out-of-sample median VaR over time

Model	$\tau = 5\%$		$\tau = 1\%$		$\tau = 0.1\%$	
	# Exc	$\hat{\tau}$ in %	# Exc	$\hat{\tau}$ in %	# Exc	$\hat{\tau}$ in %
1) Gaussian	23	5.42	6	1.42	4	0.94
2) Student-t	24	5.66	6	1.42	0	0
3) Stable	23	5.42	3	0.71	0	0
4) Nonparametric	22	5.19	4	0.94	0	0
5) Weekly GARCH(1,1)	12	2.83	3	0.71	0	0
6) Weekly FIGARCH(1,1)	11	2.59	5	1.18	0	0
7) 10-min GARCH(1,1)	18	4.25	4	0.94	1	0.24
8) 10-min FIGARCH(1,1)	21	4.95	5	1.18	1	0.24
9) Realized GARCH(1,1)	20	4.72	4	0.94	0	0
10) Realized FIGARCH(1,1)	34	8.02	6	1.42	0	0
11) FRM	10	2.36	4	0.94	0	0

Table 4.3: In-sample weekly Value at Risk exceedances with according probabilities for given confidence levels 95%, 99% and 99.9%

with 95% confidence bounds for confidence levels 95% (green), 99% (red) and 99.9% (blue). Whereas the Stable models holds the unconditional coverage by exhibiting the largest VaR of all tested models, the weekly FIGARCH models ensures coverage by time-varying VaR forecasts, resulting, on average, in smaller VaR forecasts. Although VaR forecasts from the frequency rescaling method (FRM) are comparable to the weekly FIGARCH models in terms of size, the FRM has the advantage of producing time-constant VaR in-sample forecasts. Narrower confidence bounds for the VaR forecasts are supported out-of-sample. Out of the VaR models presented, the FRM is presented to be highly beneficial for institutional investors dealing with lower sampling frequencies such as weekly, quarterly and annual. Holding the unconditional coverage at low values of VaR, in contrast to the Lévy stable motion, the VaR forecasts are not time-dependent as in the GARCH

models. Moreover portfolio balance sheets should not only report VaR levels, but also hedge respective risks, which reduces portfolio turnovers and transaction costs. In comparison to the nonparametric VaR estimation from weekly data, the FRM utilizes a semiparametric estimation procedure to scale high-frequency data to the lower frequency, inducing tail events, which never happened in the original data history.

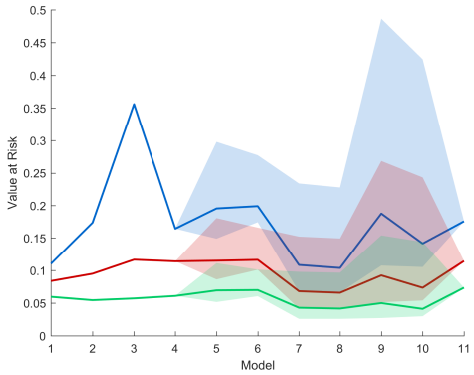

Figure 4.18: In-sample median Value at Risk over time with 95% confidence bounds for confidence levels 95% (green), 99% (red) and 99.9% (blue)

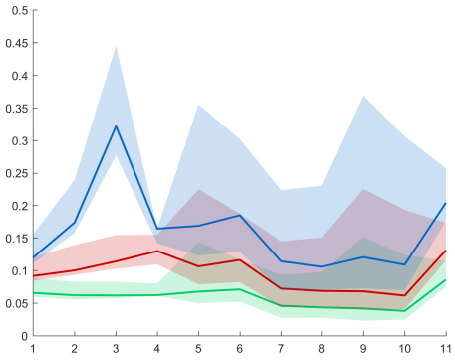

Figure 4.19: Out-of-sample median Value at Risk over time with 95% confidence bounds for confidence levels 95% (green), 99% (red) and 99.9% (blue)

4.6 Conclusion

Estimating risk in a low-frequency context and limited data is difficult for large confidence levels as relevant weekly, quarterly or annual data capture an insufficient history or are not relevant for the risk profile of the financial institution. Utilizing high-frequency data for the estimation of low-frequency risk measures can provide significantly more data points, given that a fractal structure could be verified among different sampling frequencies. The proposed method incorporates

Model	$\tau = 5\%$		$\tau = 1\%$		$\tau = 0.1\%$	
	# Exc	$\hat{\tau}$ in %	# Exc	$\hat{\tau}$ in %	# Exc	$\hat{\tau}$ in %
1) Gaussian	11	2.59	6	1.42	4	0.94
2) Student-t	13	3.07	6	1.42	0	0
3) Stable	14	3.30	4	0.94	0	0
4) Nonparametric	15	3.54	4	0.94	2	0.47
5) Weekly GARCH(1,1)	10	2.36	4	0.94	1	0.24
6) Weekly FIGARCH(1,1)	13	3.07	4	0.94	0	0
7) 10-min GARCH(1,1)	28	6.60	7	1.65	3	0.71
8) 10-min FIGARCH(1,1)	30	7.08	8	1.89	4	0.94
9) Realized GARCH(1,1)	36	8.49	16	3.77	4	0.94
10) Realized FIGARCH(1,1)	37	8.73	14	3.30	4	0.94
11) FRM	7	1.65	3	0.71	0	0

Table 4.4: Out-of-sample weekly Value at Risk exceedances with according probabilities for given confidence levels 95%, 99% and 99.9%

a specific form of rescaling, involving filters for seasonality and time dependence in the α-stable framework. In contrast, given finite variance and the classic CLT, Gaussianity cannot be circumvented for long holding periods, which is not supported empirically. In an i.i.d. simulation study we indicate that ten years of data barely suffice in order to efficiently estimate weekly VaR of a stock market portfolio for a confidence level of 99.9% and larger. For the relevant confidence levels, the authors recommend to use at least 10^3 observations for the 99% confidence, 10^4 observations for the 99.9% confidence and 10^5 observations for the 99.99% confidence level. As deficiencies through underestimation of VaR hold for relevant i.i.d. distributions such as Student-t or Stable, the predicament unfolds in aggravated form for Expected Shortfall under time-dependent processes. The high-frequency data-set of NASDAQ portfolio returns gives evidence for dif-

ferent generating distributions over the sampling frequencies. In order to obtain events for the lower sampling frequencies, we show that by filtering high-frequency returns for seasonality and long-range dependence, the mesofractal assumption (Lévy stable motion) cannot be denied for the residuals of sampling frequencies larger than five minutes. Within the α-stable framework and the underlying fractal structure, 5-min filtered log-returns are scaled to the weekly minimum liquidity horizon. The obtained weekly data points cover tail events which never existed in the original data history, which improve the estimation of the respective risk measures. Empirically, the backtest reveals that the frequency rescaling method holds the unconditional coverage or all confidence levels reported, in- and out-of sample. Given all models which hold the unconditional coverage, the VaR from the frequency rescaling method are the smallest over time, involving no large deviations given the underlying iid assumptions for the weekly frequency. The open question remains to be if the data series we observe are just preasymptotic snapshots of a fractional Lévy stable motion (Samorodnitsky and Taqqu, 1994). For the case that there are processes, which involve both properties over different scales, the only relevant filter for the higher frequencies remains to be microstructure noise and seasonality, which are generated due to market imperfections and cyclical business components. If so, the diverging behavior within the sampling frequencies of the process is merely a relic of lacking data points in the lower sampling frequencies, which we indicated in this paper.

Chapter 5

Utilizing high-dimensional high-frequency data for lower sampling frequencies (Multivariate)

5.1 Introduction

In order to model the whole stock market distribution properly plenty of observations are necessary. As high-frequency information are more readily available, they can provide a promising foundation for modeling, reflecting a manifold of observations compared to low-frequency data. But investors, regulators and other financial stakeholders are rarely interested in high-frequency data and their according statistics, implying daily or higher frequencies. To large extents, the interest lies in weekly, monthly or yearly figures for performance, risk or stability (BIS, 1996). The relevant figures, especially estimates for risk depending on the tail of the relevant (loss) distributions, cannot be sufficiently estimated, as the data history is limited for the respective sampling frequencies or the whole data history is not relevant for the profile of the object of interest. The transition of

high-frequency information to lower frequencies, especially related to risk, is quite often done under the Gaussian assumption implying the $T^{1/2}$-rule (Gaussian Scaling). Allowing for heavy tails under the α-stable assumption, the $T^{1/2}$-rule can be generalized to the $T^{1/\alpha}$-rule (Stable scaling). We show that both scaling regimes are not sufficient to model the tails of stock market return distributions. Subsequently, we introduce the Frequency Rescaling Method (FRM), which rescales a high-frequency data distribution to the lower sampling frequency by accounting for intraday seasonality, long range dependence and heavy tails in different sampling frequency domains. Accordingly we provide a distribution for the lower sampling frequency, which is more heavy tailed than under Gaussianity and the $T^{1/2}$-rule, but more platykurtic than the $T^{1/\alpha}$-rule.

5.2 Model

The price of an asset for a fixed horizon $T \in \mathbb{N}^+$ is given as

$$S_T = S_0 \prod_{t=1}^{T} (1 + \tilde{X}_t), \;\; \tilde{X}_t = S_t / S_{t-1} - 1, \tag{5.1}$$

where $\tilde{X}_t \sim \tilde{\mathrm{F}}$ are discrete returns in time t coming from an unknown distribution $\tilde{\mathrm{F}}$, representing price changes in per cent from period t to $t-1$. As the logarithm of a product is the sum of the logarithms, we avoid multiplicative versions of the CLT by

$$S_T = S_0 \prod_{t=1}^{T} \exp(X_t), \;\; X_t = \log\left(S_t / S_{t-1}\right) \tag{5.2}$$

$$= S_0 \exp\left(\sum_{t=1}^{T} X_t\right) = S_0 \exp(X_T), \;\; X_T \stackrel{\text{def}}{=} \sum_{t=1}^{T} X_t, \tag{5.3}$$

where $X_t \sim \mathrm{F}$ are log returns in time t coming from an unknown distribution F, representing log price changes from period t to $t-1$. We are interested in modeling

the terminal distribution by the summation of higher-frequency random variables. Thus, we are convoluting the high-frequency distributions to the lower sampling frequency, which involves a CLT asymptotically.

5.2.1 CLTs

Let random variable X_t have expectation $\mu_t = \mathsf{E}(X_t)$ and variance $\sigma_t^2 = \mathsf{E}\left[\{X_t - \mathsf{E}(X_t)\}^2\right]$, which have to be finite

$$\mu_t < \infty,\ \sigma_t^2 < \infty. \tag{5.4}$$

Accordingly, the following CLT holds

$$\sum_{t=1}^{T} X_t \xrightarrow{\mathcal{L}} \mathrm{N}\left(\sum_{t=1}^{T} \mu_t, \sum_{t=1}^{T} \sigma_t^2\right), \tag{5.5}$$

which implies that

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} (X_t - \mu_t) \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, \sigma^2\right),\ \sigma^2 = \sum_{t=1}^{T} \sigma_t^2. \tag{5.6}$$

If the distribution in horizon T is modelled as the convolution of higher frequency distributions, the process of returns, which may not be Gaussian, but of finite variance, converges to the Gaussian. Even if the process X_t is stationary weakly dependent (short memory), implying

$$\sum_{l=1}^{\infty} |\gamma(l)| < \infty \tag{5.7}$$

with auto-covariances $\gamma(l) = \mathrm{Cov}\left\{X_l, X_0\right\},\ l \in \mathbb{Z}$, the asymptotic distribution is Gaussian,

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} (X_t - \mu_t) \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, \sigma^2\right) \tag{5.8}$$

where $\sigma^2 = \sum_{l=-\infty}^{\infty} \gamma(l)$ is the long-run variance (Bai et al., 2016). We leave the domain of attraction of Gaussianity if we loose Assumptions 5.4. Presuming that the process X_t is heavy-tailed (Bai et al., 2016)

$$\mathrm{P}(X_t) > x \sim A\frac{1+\beta}{2}x^{-\alpha}, \ \mathrm{P}(X_t) < -x \sim A\frac{1-\beta}{2}x^{-\alpha} \tag{5.9}$$

for $x \to \infty$ with constant $A > 0$, $\beta \in [-1, 1]$ and $\alpha \in (1, 2)$ implying $\mu_t < \infty$ and

$$\sigma^2 = \infty. \tag{5.10}$$

The infinite moment assumption implies that the standard CLT of Equation 5.6 does not hold anymore as Gnedenko and Kolmogorov (1954a) show that

$$T^{-\frac{1}{\alpha}} \sum_{t=1}^{T} (X_t - \delta_t) \xrightarrow{\mathcal{L}} \mathrm{S}(\alpha, \beta, \gamma, 0), \tag{5.11}$$

where $\mathrm{S}(\alpha, \beta, \gamma, 0)$ is a α-stable random variable with stable exponent α, skewness β, scale γ and location δ. Losing Assumption 5.7 by introducing X_t as strongly dependent (long memory) with auto-covariance

$$\gamma(l) \sim c_\gamma l^{2H-2}, \tag{5.12}$$

Hurst exponent $H \in (0.5, 1)$, leads to limit theorem (Dobrushin and Major, 1979; Taqqu, 1979)

$$\frac{1}{n^H} \sum_{t=1}^{T} (X_t - \mu_t) \xrightarrow{\mathcal{L}} cZ_{m,H}. \tag{5.13}$$

c is a function of c_γ. $Z_{m,H}$ is a Hermite process with $m \in \mathbb{N}^+$. The Hermite process can be expressed as multiple Wiener-Itô integral, non-Gaussian for $m \geq 2$.

5.3 Data

The data-set from Lobster contains 11 years of transaction level data (Level 2) from 2007-06-27 till 2018-05-27 for the fourteen biggest NASDAQ 100 stocks by

market capitalization. Whereas the price series are depicted on Figure 5.1 (left), Figure 5.1 (right) visualizes the Apple mid price for different sample frequencies, which we aim to analyse coherently.

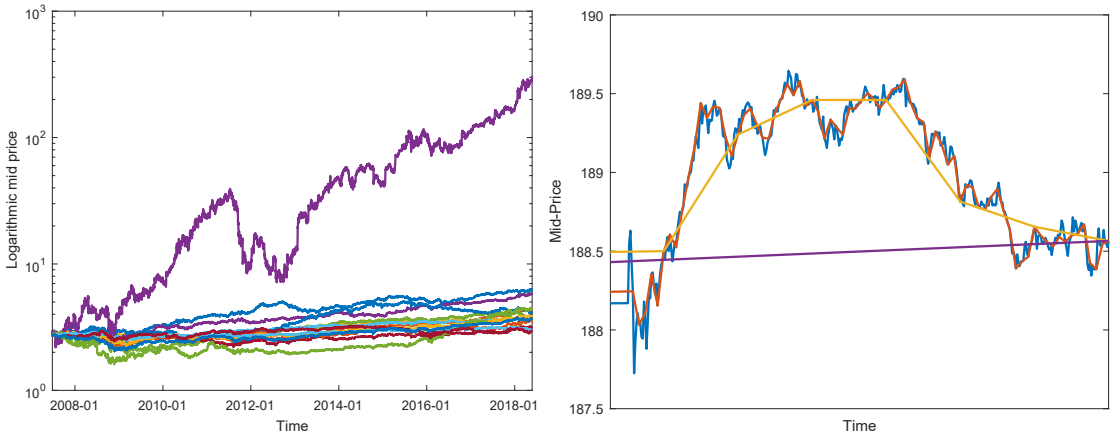

Figure 5.1: Normalized mid price data for NASDAQ stocks 2007-06-27 till 2018-05-27 (left) and the last day of Apple mid price data for 1-minute (blue), 5-minute (red), 1-hour (orange) and 1-day (violet) frequency (right)

An important consequence of sampling data at different frequencies for the same time span are sample size effects. Namely, given 11 years of data, we obtain increasingly more data points for higher sampling frequencies (see Table 5.1).

Sampling frequency	1 min	1 hour	1 day	1 week
Number of log returns	1071719	17861	2747	548

Table 5.1 Data points over sampling frequencies

5.4 Estimation

5.4.1 Mesofractality

Whereas the semi-log densities of 5-min log returns (see Figure 5.2, left) indicate heavy tails far beyond the Gaussian, the densities for weekly log returns (see Figure 5.2, right) are not as heavy tailed, still differing significantly from the Gaussian assumption. Specifically, the tails of the weekly distributions are not containing as many non-normal events, which might be due to sample-size effects.

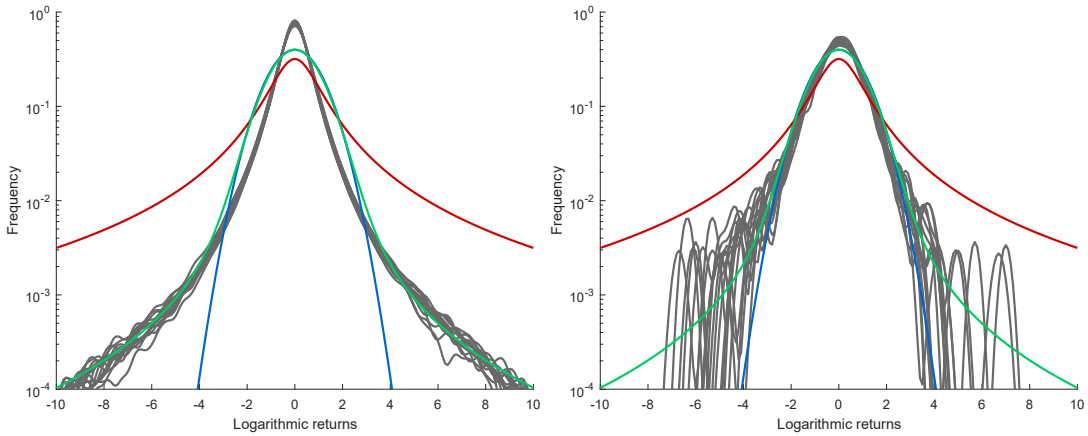

Figure 5.2: Semi-log kernel densities of 5-min (left) and weekly (right) NASDAQ log-returns (gray) under Gaussian ($\alpha = 2$) (blue), Stable ($\alpha = 1.8$) (green), Cauchy ($\alpha = 1$) (red)

We argue centrally, first in Wesselhöfft and Härdle (2019), by using a block-bootstrap procedure for returns of different sampling frequencies, that the observed convergence to Gaussianity is solely due to sample-size effects. Accordingly, we draw blocks of the size of the weekly sampling frequency. Whereas the sample kurtosis for all NASDAQ stocks decreases exponentially with sample size (see Figure 5.3, right), we show in Figure 5.3 (left), that the block-bootstrapped Apple kurtosis estimators (red) are approximately constant over sampling frequencies. The result holds for all NASDAQ assets analyzed. The indication is subsequently,

that if we would observe the same amount of observations for the low sampling frequency, the sample kurtosis will be of comparable, large size. As such, we presume that with the increasing number of observations, the sample kurtosis tends to infinity, speaking in favor of α-stable distributions as moments of α-stable r.v. are infinite in the limit for moments of order $p > \alpha$, $p \in \mathbb{R}^+$, including infinite kurtosis and also variance for all analyzed stocks.

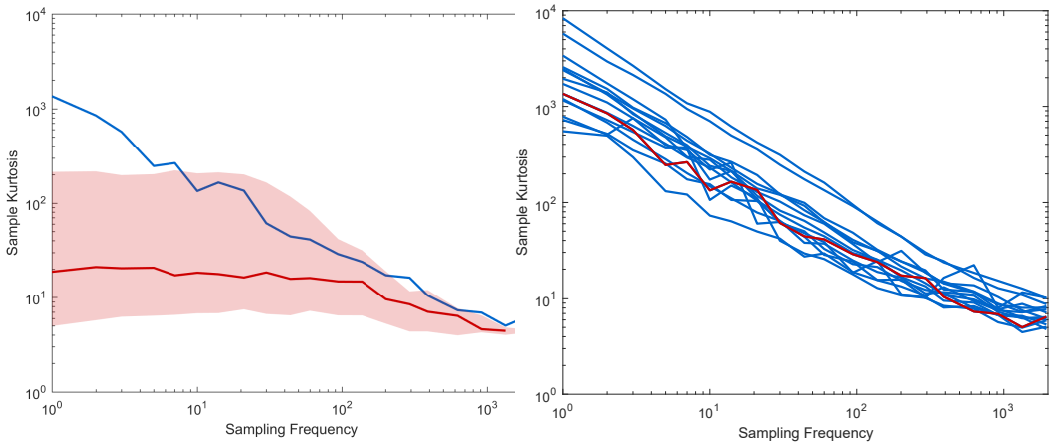

Figure 5.3: Sample kurtosis of Apple returns (blue) with bootstrapped kurtosis (red), using blocks of 548 data points representing the weekly frequency (left) and sample Kurtosis over frequency for NASDAQ stocks (blue) including Apple (red) (right)

Instead of focussing on non-converging estimators like sample kurtosis, we can estimate the α-stable exponent for different sampling frequencies, providing a bounded framework to analyze the tails of different distributions. Accordingly, Figure 5.4 (left) depicts the α-stable exponent, indicating that for all NASDAQ assets, returns get heavier tailed with higher sampling frequencies. Assessing the skewness of the distribution in Figure 5.4 (right), we find no from zero significant β over sampling frequencies and assets, which simplifies the modeling in the multivariate framework.

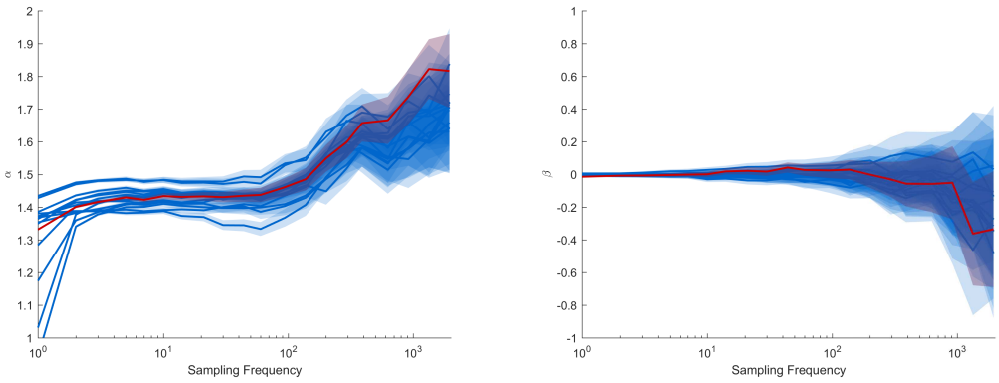

Figure 5.4: Stable parameters α (left) and β (right) including Apple (red) over sampling frequencies, 95% confidence intervals from the MLE

5.4.2 Linear dependence through elliptically stable distributions

Starting to assess the dependence structure of the NASDAQ log returns over sampling frequencies empirically, the linear Pearson correlation coefficient for all asset with Apple is plotted in Figure 5.5. In contrast to tail and memory behavior, we indicate that correlations are not changing significantly over samplings frequencies. In the context of multivariate modeling, α-stable distributions involve, as for the univariate case, characteristic functions. As all relevant assets are heavier tailed than under Gaussianity (Figure 5.4, left), implying $\alpha_j < 2,\ j = 1, \ldots, k$. Hence, not only the variance is infinite, but also the linear Pearson correlation coefficient

$$\lim_{\sigma_i, \sigma_j \to \infty} \rho_{i,j} = \frac{\mathrm{Cov}(X_i, X_j)}{\sigma_i \sigma_j} = 0,\ \ i, j = 1, \ldots, k \tag{5.14}$$

tends to zero. Serving as a natural, symmetric generalization of the multivariate Gaussian, elliptically contoured α-stable laws are introduced, which can be efficiently estimated for dimensions $k \leq 40$ (Nolan, 2013). As for the univariate case, this class of distributions has its own domain of attraction, implying that the distribution preserves its shape under aggregation in the presence of linear de-

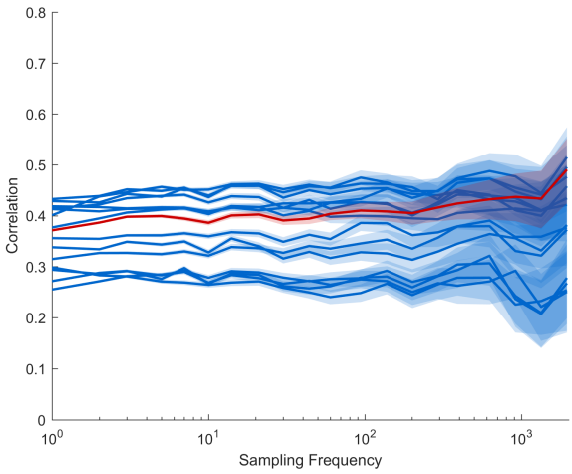

Figure 5.5: Pearson correlations with Apple over assets and sampling frequencies including Apple and Amazon (red) with 95% confidence bounds estimated from the MLE

pendence. Random vector $X \sim \mathrm{S}_k(\delta, \Gamma, \psi)$ is elliptically distributed with positive definite scaling matrix $\Gamma = AA^\top$, $A \in \mathbb{R}^{k \times k}$ and location vector $\delta \in \mathbb{R}^k$ when

$$X \overset{\mathcal{L}}{=} \delta + AY, \tag{5.15}$$

where Y is spherical with characteristic generator ψ. The characteristic function is given by

$$\varphi_X(u) = \mathsf{E}\left\{\exp\left(iu^\top X\right)\right\} = \exp\left(iu^\top \delta\right) \psi\left(u^\top \Gamma u\right). \tag{5.16}$$

A subclass of elliptical distributions are normal variance mixtures $X = [X_1, \ldots, X_k]^\top$ for

$$X \overset{\mathcal{L}}{=} W^{1/2} AZ + \delta, \tag{5.17}$$

with $Z \sim \mathrm{N}(0, I_k)$ being k-dimensional Gaussian and $W \geq 0$ being a non-negative one-dimensional random variable, independent of Z (Kring et al., 2009).
A further subclass of normal variance mixtures are α-stable sub-Gaussian $X = [X_1, \ldots, X_k]^\top$ for $W \sim \mathrm{S}(\alpha/2, (\cos \pi\alpha/4)^{2/\alpha}, 1, 0)$, $0 < \alpha < 2$, being α-stable distributed, parameterized following Nolan (2017). $G \sim \mathrm{N}(0, \Gamma)$ is multidimensional

Gaussian with scaling matrix $\Gamma = AA^\top$. Then $X \sim \mathrm{S}_k(\alpha, \beta, \Gamma, \delta, \psi)$, $\beta = 0$ is α-stable sub-Gaussian if

$$\begin{aligned} X &\overset{\mathcal{L}}{=} W^{1/2}G + \delta \\ &\overset{\mathcal{L}}{=} W^{1/2}AZ + \delta, \ Z \sim \mathrm{N}(0, I_k) \\ &\overset{\text{def}}{=} AY + \delta, \end{aligned} \tag{5.18}$$

while $Y \sim \mathrm{S}_k(\alpha, 0, I_k, 0)$ is radially symmetric α-stable. The according characteristic function of X is

$$\begin{aligned} \varphi_X(u) &= \int_{-\infty}^{\infty} f_X(x) \exp(iu^\top X) \mathrm{d}x = \mathsf{E}\left(iu^\top X\right) \\ &= \exp\left\{ -\left(\frac{1}{2}u^\top \Gamma u\right)^{\alpha/2} + iu^\top \delta \right\}, \end{aligned} \tag{5.19}$$

$f_X(x)$ as probability density function. $\Gamma \in \mathbb{R}^{k \times k}$ is the positive definite scale matrix and $\delta \in \mathbb{R}^k$ the location vector. The characteristic generator is therefore given by

$$\psi(s, \alpha) = \exp\left\{ -\left(\frac{1}{2}s\right)^{2/\alpha} \right\}. \tag{5.20}$$

This implies that α-stable sub-Gaussian distributions are scale mixtures of multivariate normal distributions (Samorodnitsky and Taqqu, 1994). For $\alpha = 2$, the characteristic function collapses to the Gaussian. For $G \sim \mathrm{N}(0, I_k)$, the characteristic function of Y in Equation 5.18 simplifies to

$$\varphi_Y(u) = \mathsf{E}\left(iu^\top Y\right) = \exp\left(-\gamma^\alpha |u|^\alpha\right). \tag{5.21}$$

For the lower sampling frequencies, involving T convolutions, the estimated higher sampling frequency log-returns are summed to the chosen frequency:

$$\begin{aligned} X_T = TX_t, \ X_t &\sim \mathrm{S}_k(\alpha, 0, \Gamma, \delta, \psi) \\ X_T \ &\sim \mathrm{S}_k(\alpha, 0, T\Gamma, T\delta, \psi). \end{aligned} \tag{5.22}$$

Employing the GCLT of Equation 5.11 in order to scale the high-frequency distributions to the low-frequency, requires a constant α-stable exponent over sampling

frequencies. We model the increase of α-stable exponent with the sampling frequency, as we have depicted in Figure 5.4 (left), by showing that the changes are due to varying memory patterns (long memory) in different sampling frequencies.

5.4.3 Long memory through Multifractality

Assessing the autocorrelation functions of absolute returns for different sampling frequencies in Figure 5.6 indicates different memory decaying patterns. In higher frequencies, i.e. 5-minute and 1-hour, intraday seasonalities lead to non-standard patterns, which cannot be modelled without filters intraday seasonality. For lower sampling frequencies, the long memory behavior cannot be discovered without doubt. Thus, we implement a filter for intraday seasonalities, which will lead to a clear autocorrelation pattern.

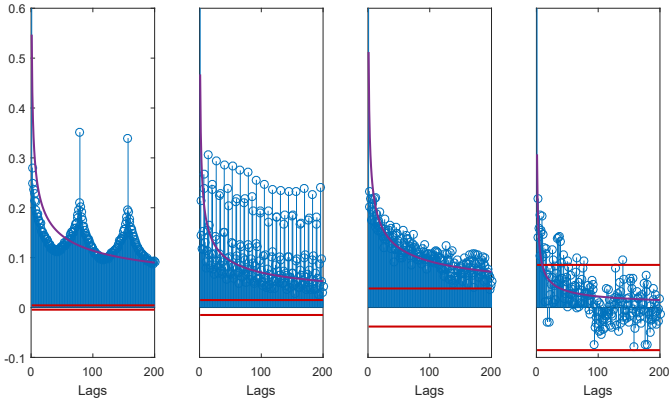

Figure 5.6: Sample autocorrelation function of absolute Apple returns for 5-minute, hourly, daily and weekly frequencies with confidence bounds (red) and hyperbolic fit (violet)

5.4.4 Frequency Rescaling Method (FRM)

We are interested in the data-generating process for stock market returns in a stochastic framework. Accordingly, there is a variety of potential models, which

can be separated into two dimensions: tail behavior and memory (see Table 5.2). High-frequency returns exhibit heavy tails and long memory, but the convolution of the densities of the increments of the processes gives low-frequency returns such as weekly, which are still heavy tailed, but are not exhibiting long memory anymore. The idea of the Frequency Rescaling Method (FRM) is to scale an altered high-frequency data set to the lower sampling frequency to provide more observations.

IID	Short Memory	Long Memory
Gaussian	GARCH	Fractional Brownian Motion, FIGARCH
Heavy tailed, Finite variance	Heavy tailed GARCH	Multifractal, heavy tailed FIGARCH
α-stable, Infinite Variance, Mesofractal	α-stable GARCH	Linear Fractional Stable Motion, α-stable FIGARCH

Table 5.2 Stochastic processes for different environments

Plotting the absolute returns of Apple returns (1-minute frequency) with average (red) and polynomial of order seven (blue) over the course of the day from 09:30 a.m to 04:00 p.m. (see Figure 5.7, left), reveals a parabolic relationship with higher absolute returns in the beginning and the end of the day. Following Engle and Sokalska (2012), we estimate the seasonal component s_m for each minute of the day $m = 1, \ldots, 390$ by

$$s_m = \frac{1}{T} \sum_{t=1}^{T} |X_{t,m}|. \tag{5.23}$$

After normalizing the raw one-minute returns by

$$Z_{t,m} = \frac{X_{t,m}}{\sqrt{s_m}}, \tag{5.24}$$

the lower frequencies are calculated. The filtered absolute one minute Apple returns, not containing the intraday seasonality anymore, are plotted over the course of the day in Figure 5.7 (right). The ACFs of the deseasonalized absolute returns in Figure 5.8 indicate long memory as the decay is hyperbolic and not exponential for frequencies daily and higher. For the weekly sampling frequency no significant autocorrelations beyond the first lags can be observed.

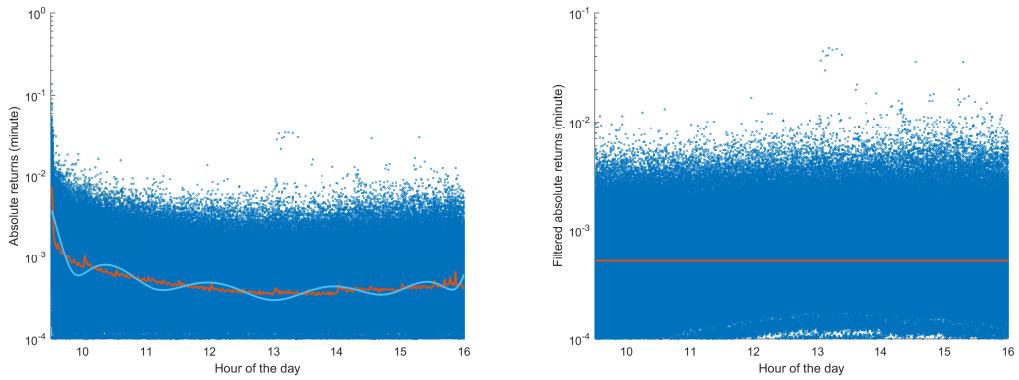

Figure 5.7: Absolute Apple returns (1-minute frequency) with average (red) over hour of the day and polynomial fit (blue) (left) and filtered Apple returns (right)

Accordingly, we model the observed phenomena of heavy tails, seasonality and long-range dependence by employing the FIGARCH methodology (Baillie et al., 1996) for seasonally filtered returns under stable laws (Paolella et al., 2002) in order to explain the instability in distributions. The Stable paretian power GARCH process $\mathrm{S}^{h}_{\alpha,\beta}\,\mathrm{GARCH}(r,s)$ with $r, s \in \mathbb{N}^{+}$, $h \in \mathbb{R}$ and seasonality component s_k is

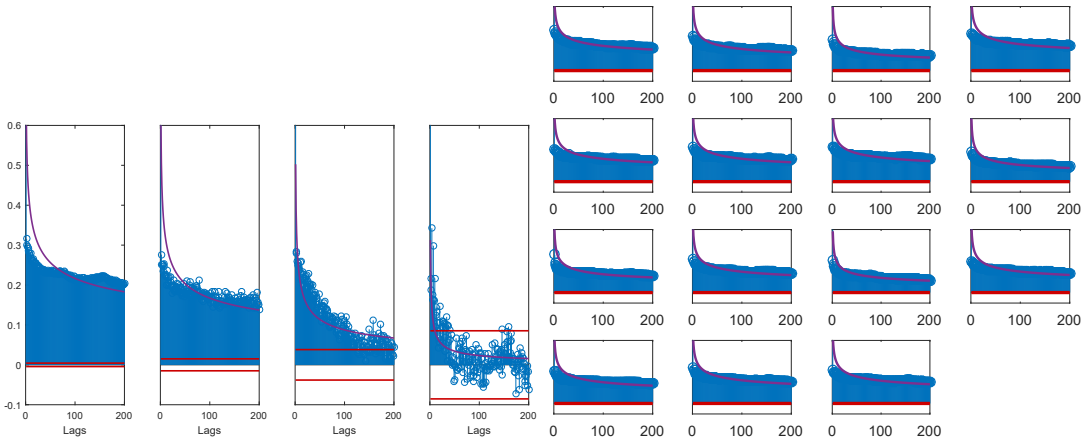

Figure 5.8: Sample ACF of deseasonalized Apple returns over frequencies (left) and all NASDAQ assets for the 5-minute frequency (right)

given by

$$X_t = \delta + s_k \gamma_t \varepsilon_t, \;\; \varepsilon_t \sim \mathrm{S}_{\alpha,\beta}(1,0) \tag{5.25}$$

$$\gamma_t^h = \theta_0 + \sum_{i=1}^{r} \theta_i |\varepsilon_{t-i}|^h + \sum_{j=1}^{s} \phi_j \gamma_{t-j}^h \tag{5.26}$$

$$= \theta_0 + \theta(L)\varepsilon_t^h + \phi(L)\gamma_t^h. \tag{5.27}$$

The according GARCH equation can be rewritten in lag polynomial form:

$$\{1 - \theta(L) - \phi(L)\}\, \varepsilon_t^h = \theta_0 + \{1 - \phi(L)\}\, \left\{\varepsilon_t^h - \gamma_t^h\right\}. \tag{5.28}$$

In order to allow for a slower decay than exponential, the fractional difference operator $(1-L)^d,\; 0 < d < 1$ is introduced, obtaining the fractionally integrated GARCH (FIGARCH) equation:

$$\{1 - \theta(L) - \phi(L)\}\, (1-L)^d \varepsilon_t^h = \theta_0 + \{1 - \phi(L)\}\, \left\{\varepsilon_t^h - \gamma_t^h\right\}. \tag{5.29}$$

In contrast to the special cases of GARCH ($d = 0$) and IGARCH ($d = 1$), Davidson (2004) shows that this class of processes is able to reproduce more flexible temporal dependencies, i.e. long memory. As GARCH processes are modelled

separately for each sampling frequency Mandelbrot et al. (1997) argues that this family of fractionally integrated models is neither self-affine nor scale consistent. Still, Fisher et al. (1997) find evidence that the class of FIGARCH-models can mimic multifractality.

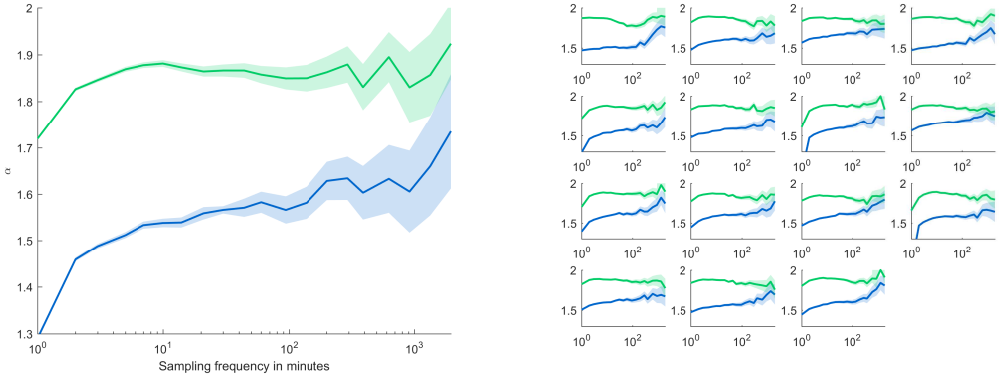

Figure 5.9: Apple (left) and NASDAQ stocks (right) stability indices of innovations over sampling frequencies with (green) and without (blue) FIGARCH filter

In Figure 5.4 (left) we observed, that the α-stable exponent increases with sampling frequency over all NASDAQ assets. Selecting Apple exemplary, see Figure 5.9 (left), the stability indices before FIGARCH filter (blue) depict the same behavior. After applying the FIGARCH filter, for sampling frequencies larger than 5-minutes, mesofractal behavior can be detected as the stability index is not varying significantly over sampling frequencies anymore. This allows to rescale the FIGARCH model residuals from high frequencies to low frequencies utilizing the Generalized Central Limit Theorem (GCLT) from Equation 5.11.

5.4.5 Frequency Rescaling Method: Algorithm

Let $X_t \in \mathbb{R}^k$ be high-frequency returns, whereas $X_T \in \mathbb{R}^k$ are low-frequency, which we aim to estimate. In the framework of elliptically stable distributions,

$$X_t \sim \mathrm{S}_k(\alpha_t, \beta_t, \Gamma_t, \delta_t) \tag{5.30}$$

$$X_T \sim \mathrm{S}_k(\alpha_T, \beta_T, \Gamma_T, \delta_T) \tag{5.31}$$

evaluate if $\alpha_t = \alpha_T$ in order to check for mesofractality. If multifractal behavior can be detected, apply the Filter 5.25 of the form $\Gamma_t = f(\varepsilon^h_{t-i}, \Gamma^h_{t-i})$, $i \in \mathbb{N}^+$ in order to obtain the high-frequency model residuals ε_t

$$\varepsilon_t \sim \mathrm{S}_k(\alpha_\varepsilon, \beta_\varepsilon, \Gamma_\varepsilon, \delta_\varepsilon), \tag{5.32}$$

which should be mesofractal, implying that the α-stable exponent is not varying over the sampling frequencies. Then, normalize ε_t by

$$Z = \Gamma_\varepsilon^{-1}(\varepsilon_t - \delta) \tag{5.33}$$

$$Z \sim \mathrm{S}_k(\alpha_\varepsilon, \beta_\varepsilon, I, 0), \tag{5.34}$$

obtaining a k-dimensional radially symmetric stable distribution, which would be standard normal for $\alpha = 2$ and $\beta = 0$. At last, rescale the radially symmetric distribution to the lower sampling frequency, as elliptically stable distributions are also closed under convolution, obeying additionally their own domain of attraction.

$$X_{\mathrm{FRM}} = \delta_{\mathrm{FRM}} + \Gamma_{\mathrm{FRM}}^{1/\alpha_t} Z \tag{5.35}$$

$$X_{\mathrm{FRM}} \sim \mathrm{S}_k(\alpha_\varepsilon, \beta_\varepsilon, \Gamma_{\mathrm{FRM}}, \delta_{\mathrm{FRM}}), \tag{5.36}$$

where $\delta_{\mathrm{FRM}} = \delta_t$ and $\Gamma_{\mathrm{FRM}} = T\Gamma_t$. Subsequently, we will denote $X_{\mathrm{FRM}} \sim \mathrm{F}_{\mathrm{FRM}}$ for the marginals. One marginal, the NASDAQ portfolio, is visualized via KDEs in Figure 5.10. The FRM allows us to rescale an altered empirical high-frequency distribution utilizing the parametric structure of elliptically stable laws beneath.

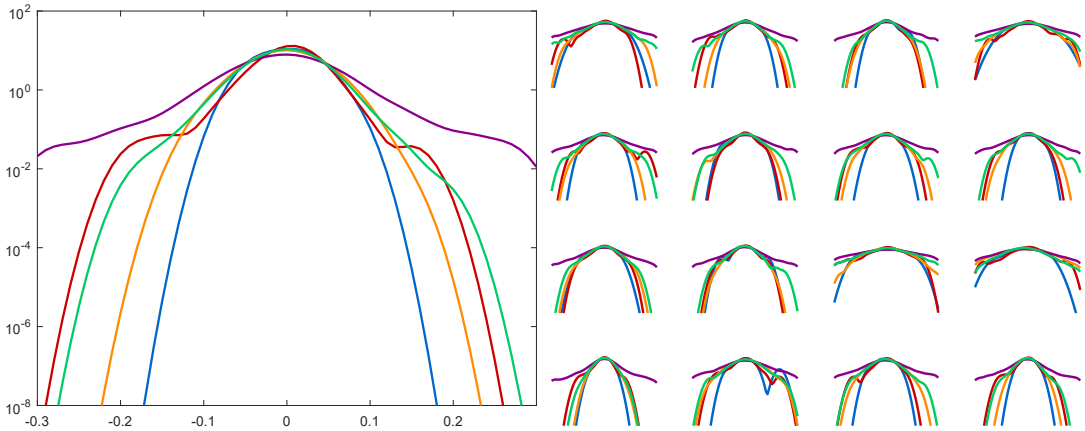

Figure 5.10: Distribution forecasts $F_{10\%}$ (green for FRM, blue for Gaussian scaling, orange for Empirical scaling and violet for stable scaling) for $F_{100\%}$(red) for NASDAQ equally distributed portfolio (left) and NASDAQ stocks (right)

5.5 Distribution forecast

In order to test the superiority of the Frequency Rescaling Method (FRM), we evaluate if the distance between the forecasted and the realized distribution is the smallest among all model competitors. As the data-set covers the time span from 2007-06-27 till 2018-05-27, we use the first 10% of the time series as training data in order to evaluate, if the FRM was able to account for the drawdowns halving the NASDAQ 100, which were starting mid August 2008.

5.5.1 Nonparametric statistical test: Kolmogorov-Smirnov

Given vectors $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^n$ with elements $x_i, y_i,\ i = 1, \ldots, n$, the Minkowski distance for $p \geq 1$ is defined as

$$d(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}, \tag{5.37}$$

collapsing to the Chebychev distance for $p \to \infty$.

$$d_{\text{Chebychev}}(x, y) = \max(|x_i - y_i|) \tag{5.38}$$

Presume that $X \sim \mathrm{F}_X$, $Y \sim \mathrm{F}_Y$ are random variables from distribution F, the distance measure is the total variation (TV) distance

$$d_{\mathrm{TV}}(x, y) = \sup(|F_X(x) - F_Y(x)|). \tag{5.39}$$

Based on the TV distance, the two-sample Kolmogorov-Smirnov statistic for n draws from F_X and m draws from F_Y (Kolmogorov, 1933; Smirnov, 1936) is constructed as

$$d_{n,m} = \sup|F_X(x) - F_Y(y)| \tag{5.40}$$

in order to test under the null H_0

$$H_0 : F_X = F_Y, \tag{5.41}$$

if the null H_0 can be denied.

$$H_1 : F_X \neq F_Y \tag{5.42}$$

We can deny the null hypothesis, if the test statistic

$$d_{n,m} > K_\alpha \left(\frac{n+m}{nm}\right)^{1/2}, \ 0 \leq \alpha \leq 1, \tag{5.43}$$

where K_α is the critical value from the Kolmogorov distribution. Accordingly, we cannot deny the null hypothesis of equality in distributions if the test statistic is larger than the respective critical value.

5.5.2 Forecasting comparison

Plotting the weekly data series of the NASDAQ equally distributed portfolio for the training sample as kernel density (Figure 5.11) suggests that the Gaussian distribution might be a sufficient fit for the data. Given knowledge of the full data-set, including the financial crisis, the Gaussian assumption has to be denied

(see Figure 5.12). As individual stock data provide the same results, the according visualizations are omitted.

Accordingly, we compare the distributions coming from the the estimated low-frequency distributions $F_s,\ 0 \leq s \leq 1$ using $s = 0.1$ of the data (10%) with the empirical realized distribution F_1. The list of model competitors is provided in Table 5.3, which are introduced via their CLTs in Subsection 5.2.1.

Model	Notation	Scaling equation
Gaussian Scaling	F_s^{Gauss}	$X_T^{\text{Gauss}} = T\mu_t + T^{\frac{1}{2}}\sigma_t Z,\ Z^{\text{Gauss}} \sim \text{N}$
Stable Scaling	F_s^{Stable}	$X_T^{\text{Stable}} = T\delta_t + T^{\frac{1}{\alpha}}\gamma_t Z,\ Z^{\text{Stable}} \sim \text{S}$
Empirical Scaling	$F_s^{\text{Empirical}}$	$X_T^{\text{Empirical}} = T\mu_t + T^{\frac{1}{2}}\sigma_t Z,\ Z^{\text{Empirical}} \sim \text{F}$
Frequency Scaling	F_s^{FRM}	$X_T^{\text{FRM}} = T\delta_t + T^{\frac{1}{\alpha}}\gamma_t \varepsilon_t^{\text{FRM}},\ \varepsilon_t^{\text{FRM}} \sim \text{S}$

Table 5.3 Distribution forecasts $F_s,\ 0 \leq s \leq 1$ Competitors against the whole empirical distribution $F_1^{\text{Empirical}}$

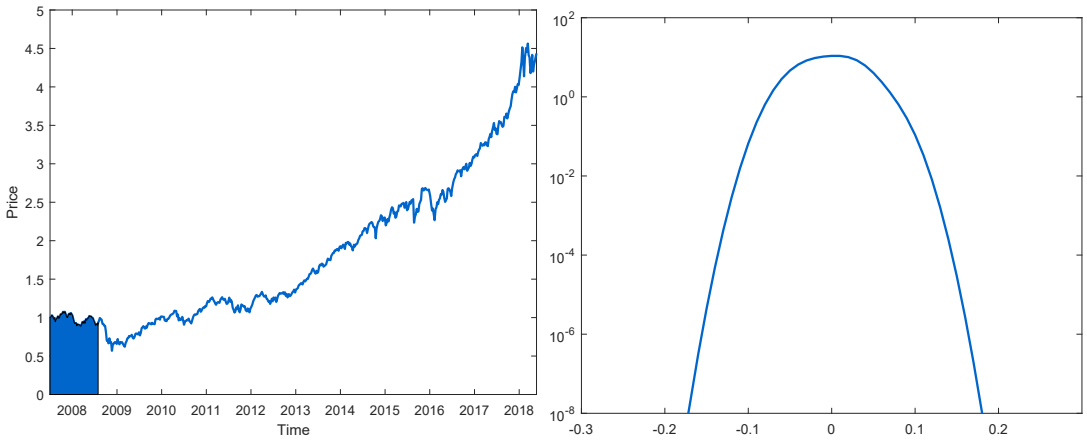

Figure 5.11: Weekly prices (left) and KDE $F_{10\%}^{\text{Empirical}}$ for weekly log returns (Training sample) (right)

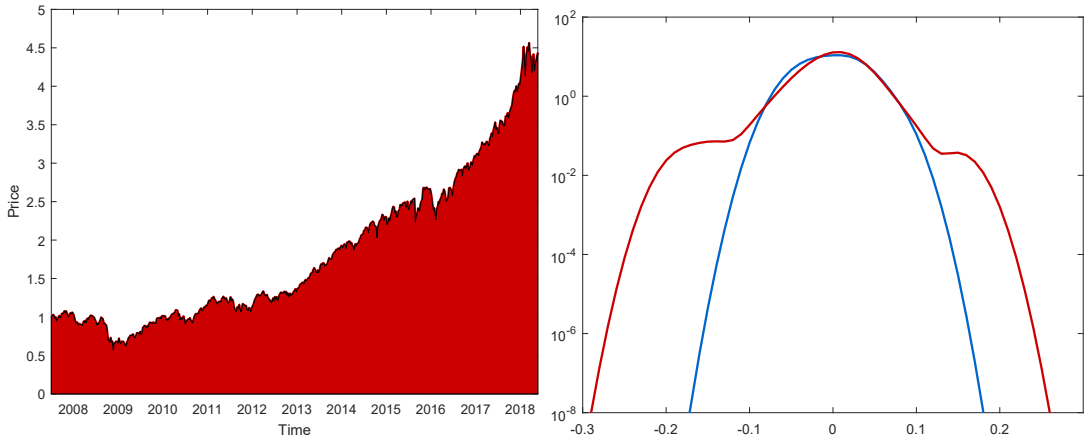

Figure 5.12: Weekly prices (left) and KDE of $F_{100\%}^{\text{Empirical}}$ (red) and $F_{10\%}^{\text{Empirical}}$ (blue) for weekly log returns (right)

In Figure 5.13 we provide the Kolmogorov-Smirnov statistics of the model competitors with increasing time span for the data, starting at $s = 0.1$ 10% up to $s = 1$ 100%. Naturally, $F_s^{\text{Empirical}} = F_1^{\text{Empirical}}$ for 100%, so the KS-Statistic is zero. But for $s = 0.1$, the KS-statistic is lowest for the FRM, generating $F_{0.1}^{FRM}$. Accordingly, a severely limited time frame does suffice for the FRM to outperform all competitors in terms of the TV-distance. Taking a look beyond the data-set by F_s, $0 \leq s \leq \infty$, where F_∞ denotes the data generating process: After the financial crises, focussing on the period starting from 2010, we observe that the KS-Statistic for the empirical weekly distribution is dominating all other distribution forecasts. But $F_1^{\text{Empirical}}$ is not the data generating process, which we aim to forecast. Given that we aim to forecast $F_s^{\text{Empirical}}$, $s \to \infty$, we could presume following the same argument, that $F_1^{\text{Empirical}}$ will not be the best forecast for $F_\infty^{\text{Empirical}}$, which will be even more heavy tailed, speaking in favor of F_1^{FRM} in order to forecast $F_\infty^{\text{Empirical}}$. The future will incorporate crises on bigger scales (low-frequency), which are not yet observable, but by utilizing the self-affine structure, already observed smaller scale (high-frequency) crises be used for the bigger scale (low-frequency) as we have shown.

Figure 5.13: KS-statistics with increasing In-Sample period for Empirical (blue), FRM (green), $T^{1/2}$ (orange) and $T^{1/\alpha}$ (violet)

5.6 Conclusion

The tails of the distribution are the core driver of the whole distribution for financial assets, especially talking about risk. But relevant financial figures are interested in scales or sampling frequencies like weekly, which relevant data series are not readily available for large time spans. Accordingly. we propose a method coined Frequency Rescaling Method (FRM), which allows to rescale a filtered high-frequency distribution to the lower sampling frequency, obtaining events which never occurred in the original data history. Employing the nonparametric Kolmogorov-Smirnov test, we show that the FRM is able to outperform the weekly NASDAQ distribution forecasts coming from the empirical data set, the $T^{1/2}$-rule and the $T^{1/\alpha}$-rule under α-stability.

Chapter 6

A statistical classification of cryptocurrencies

6.1 Introduction

Cryptocurrencies, served as a new digital asset, have attracted much attention from investors and academics. Along with this growing popularity, the market capitalization of cryptocurrencies is increasing substantially. According to a recent report (Transparency Market Research, 2018), the total capitalization for cryptocurrencies market was around US$ 574.3 mn in the year 2017 and is expected to reach US$ 6702.1 mn by the end of 2025. Most articles focus on Bitcoin (BTC), as it is considered the first decentralized cryptocurrency, which has the largest capitalization from its beginning till now. An extensive review of the literature regarding the Bitcoin can be found in Corbet et al. (2019). Appendix 6.5 lists a synthesis of the empirical findings regarding the statistical properties of cryptocurrencies, compared to classical assets.

There is no standard, widely accepted definition of cryptocurrencies allowing to identify them within the existing economic theory (Núñez et al., 2019). In general,

cryptocurrencies are defined as "digital representations of value, made possible by advances in cryptography and distributed ledger technology (DLT)" (Fund, 2019). Followig the IMF report, cryptocurrencies can be separated into two branches: Bitcoin-like crypto assets (BLCAs) and digital tokens. BLCAs are digital assets based on distributed ledger technology and designed to work as a medium of exchange (for example: Bitcoin, Ether, Ripple (XRP), Bitcoin Cash, EOS, Stellar and Litecoin) (Fund, 2019). On the other hand, digital tokens can be split in four categories, depending on their economic function: payment tokens, utility tokens, asset tokens and hybrid tokens. However, there is no classification of cryptocurrencies based on their risk profile, which may be extremely beneficial for investors. In this paper we are providing a classification of assets universe, showing that cryptocurrencies poses unique statistical features, allowing them to differentiate from classical assets. Our approach is different from the existing literature, as most of the reviewed paper are using a low-dimensional approach while trying to differentiate cryptocurrencies from classical assets. Given preceding results from the literature, our contribution to the studies dealing with cryptocurrencies market is mostly empirical, proving the complete separation of cryptocurrencies from the other assets and their synchronic evolution.

Through means of dimensionality reduction techniques (Factor Analysis), we prove that most of the variation among cryptocurrencies, stocks, exchanges rates and commodities can be explained by three factors: the tail factor, the moment factor and the memory factor. Additionally, cryptocurrencies are classified into disjoint clusters, based on their statistical properties. Our results add to the findings from literature by showing that the most important factor which differentiates cryptocurrencies from classical assets is the tail behaviour of the log-returns distribution, as proven by classification techniques: Binary Logistic Regression, Discriminant Analysis, Support Vector Machines, K-means clustering and Variance Components Split. This finding is confirmed by classical Factor Analysis, per-

formed on a static basis and also by using an expanding window approach, where the assets universe can be observed in its evolutionary dynamic. The main result of our paper is the complete separation of cryptocurrencies from classical asset types in finance, by using the Maximum Variance Components Split method and other benchmark methods, (Binary Logistic Regression, Discriminant Analysis, Support Vector Machines, K-means clustering) which provide an almost complete separation. Another important result is the discovery of synchronic evolution of cryptocurrencies, compared to classical assets types. Synchronicity refers to the fact that individual cryptocurrencies tend to develop certain similar characteristics over time that make them fully distinguishable from classical assets, i.e. they tend to behave like a homogeneous group, with certain characteristics that individualize them in the assets ecosystem. By using an expanding window approach, we are able to show that cryptocurrencies have a convergent dynamic in relationship to classical assets and this convergence is driven mainly by the tail behaviour of the log-returns distribution. Moreover, cryptocurrencies as a species exhibit a divergent evolution in relation to classical assets. A related analysis can be found in ElBahrawy et al. (2017), where the cryptocurrency market is seen as an evolutive system with several characteristics which are preserved over time. According to ElBahrawy et al. (2017), the evolution of the cryptocurrency market has been ruled by “neutral" forces, i.e. no cryptocurrency has shown any strong selective advantage over the other.

The paper is subsequently organized as follows: the second section describes the methodology used, including Factor Analysis, Binary Logistic Regression, Support Vector Machines (SVM), K-means clustering, Variance Components Split (VCS) methods and the Evolutive divergence; the third section describes the datasets and interprets the results of the classification; the fourth section describes the synchronic evolution of cryptocurrencies, while the last section concludes. The codes used to obtain the results in this paper are available via `www.quantlet.de`.

6.2 Methodology

The methodology used in this paper has four layers: first, we study the statistical properties of the daily log-returns of the selected assets and we estimate the components of a multidimensional vector describing the behaviour of the time series of assets' daily log-returns. Second, we apply data dimension reduction and orthogonalization methods (Factor Analysis) in order to retain the orthogonal factors which maximize the explained variance and could discriminate between cryptocurrencies and classical assets. Third, we employ classification techniques (Binary Logistic Regression, Discriminant analysis, Support Vector Machine, K-means clustering, Variance Components Split methods) to obtain the most influential factors discriminating between cryptocurrencies and classical assets. Fourth, we prove the validity of the synchronic evolution, showing that cryptocurrencies pose specific characteristics, allowing them to differentiate over time from classical assets.

6.2.1 Taxonomy variables

In order to properly classify the assets universe, we need an initial dataset of variables that have the statistical power to differentiate between cryptocurrencies and classical assets (stocks, exchange rates and commodities). Before introducing the multidimensional dataset used for taxonomy, we set the following notation:

- n: the number of assets in the dataset;
- t: the time index, $t \in \{1, \ldots, T\}$, where T is the time of the last record in the dataset;
- $P_{i,t}$: the closing price for asset i in day t, with $i = 1 \ldots n$, $t = 1 \ldots T$;

- $R_{i,t} = \log P_{i,t} - \log P_{i,t-1}$: the daily log-return for asset i in day t, with $i = 1 \ldots n$, $t = 1 \ldots T$;
- $R = (R_{i,t})_{i=1,\ldots,n}^{t=1,\ldots,T} \in M(n,T)$: the initial matrix of the assets' daily log-returns;
- p: the number of variables used for taxonomy.

The multidimensional dataset used for taxonomy is the matrix $X_t = (x_{it,k})_{\substack{i=1\ldots n \\ k=1\ldots p}} \in M(n,p)$, whose components are detailed below, estimated for the time interval $[1,t]$, with $t = t_0, .., T$, where $t_0 = [T/3]$ (the integer part of $T/3$). First, we took into account the central moments of the log-returns distribution, through the following parameters:

- Variance: $\sigma_{i,t}^2 = \mathsf{E}\left\{(R_i - \mu_{i,t})^2\right\}$;
- Skewness: $\text{Skewness}_{i,t} = \mathsf{E}\left\{(R_i - \mu_{i,t})^3\right\} / \sigma_{it}^3$;
- Kurtosis: $\text{Kurtosis}_{i,t} = \mathsf{E}\left\{(R_i - \mu_{i,t})^4\right\} / \sigma_{it}^4$.

Second, we estimate the following parameters of the α-stable distribution, fitted to daily log-returns, in order to capture tail dependent behavior:

- Tail exponent: *Stable* $\alpha_{i,t} \in (0,2]$, lower values indicating heavier tails;
- Scale parameter: *Stable* $\gamma_{i,t} \geq 0$.

The α-stable distributions are a well-known class of distributions used in financial modelling (Rachev and Mittnik, 2000a), capturing the fat tails and the asymmetries of the real-world log-returns distributions. The α-stable parameters were estimated using the empirical characteristic function method, following Koutrouvelis (1980, 1981), through the Matlab library *stbl* (Veillette, 2012). Third, we estimated the quantiles and the conditional tail expectations for the distribution of log-returns, in order to capture the tail behaviour:

- quantiles: $Q_{\alpha;it}$, with $\alpha \in \{0.005, 0.01, 0.025, 0.05, 0.95, 0.975, 0.99, 0.995\}$;
- conditional left tail expectation: $CTE_{\alpha,it}(R_{it}) = \mathsf{E}[R_{it}|R_{it} < Q_{\alpha;it}]$, for $\alpha \in \{0.005, 0.01, 0.025, 0.05\}$;
- conditional right tail expectation: $CTE_{\alpha,it}(R_{it}) = \mathsf{E}[R_{it}|R_{it} > Q_{\alpha;it}]$, for $\alpha \in \{0.95, 0.975, 0.99, 0.995\}$.

From a market risk perspective, the left tail quantiles can be assimilated to Value-at-Risk, the conditional left tail expectation can be regarded as Expected Shortfall, while the conditional right tail expectation can be seen as the Expected Upside. Fourth, we estimated an GARCH(1,1) model in order to capture the (G)ARCH volatility model parameters. Thus, from the following variance equation of the GARCH(1,1) model

$$\sigma_t^2 = \kappa + \theta_1 \sigma_{t-1}^2 + \omega_1 \varepsilon_{t-1}^2, \tag{6.1}$$

we estimate the GARCH parameter θ_{1it} and the ARCH parameter ω_{1it}. Our multi-dimensional dataset can be seen as a tensor $\mathscr{X} \in \mathbb{R}^{n \times p \times T'}$, where n is the number of assets, $p = 23$ is the number of variables and $T' = T - t_0$ is the number of time points.

6.2.2 Factor Analysis

The most popular methods used to synthesize and extract relevant information from large datasets are Principal Components Analysis (PCA) and Factor Analysis (FA) (Bartholomew, 2011). Factor Analysis has been extensively used in financial modeling for classification purposes: Stevens (1973) who applied this technique on mergers and acquisitions or Yoshino and Taghizadeh-Hesary (2015), who analyzed credit risks for financing small and medium-sized enterprises in Asia. In this paper,

we use Factor Analysis to extract the main factors explaining the variation in the initial dataset. PCA itself is a linear combination of variables, while FA is a measurement model of a latent variable. The aim of Factor Analysis is to explain the outcome of the p variables in the data matrix X using fewer variables, the so-called factors (Härdle and Simar, 2012). The orthogonal factor model is given by:

$$X = QF + U + \mu, \tag{6.2}$$

with the following notations: X is the initial matrix of p variables, F are the common k factors ($k << p$), Q is a matrix of the non-random loadings of the common factors F, U is a matrix of the random specific factors and μ is the vector of the means of initial p variables. It also holds that the random vectors F and U are unobservable and uncorrelated.

In our paper, the initial factor pattern is extracted using the principal component method, followed by a VARIMAX rotation to insure orthogonality of the factors. The Factor Analysis is applied on the entire dataset X_T, the p initial variables being estimated for the entire time period $[1, T]$. The p-dimensional dataset X_T is then projected on the k-dimensional space defined by the k orthogonal factors, in order to observe a separation of the components of the assets.

6.2.3 Assets Classification

In order to perform the assets classification, we are using several classification techniques: Binary Logistic Regression, Discriminant Analysis, Support Vector Machines, K-means clustering and Variance Components Split. Most of these techniques have been successfully applied in relation to cryptocurrencies and classical assets. Thus, (Fischer et al., 2019) used Binary Logistic Regression, Support Vector Machines and Random Forests classifier to forecast the evolution of

cryptocurrencies returns using minute-binned data. Mirtaheri et al. (2019) used Binary Logistic Regression to identify and characterize cryptocurrency frauds that are carried out in social media.

Binary Logistic Regression

The Binary Logistic Regression model quantifies the performance of each of the orthogonal factors extracted through the Factor Analysis to discriminate between cryptocurrencies and classical assets. Thus, we are estimating the following family of models:

$$P(Y_i = 1) = \frac{\exp(\beta_{0j} + \beta_{1j} F_{ji})}{1 + \exp(\beta_{0k} + \beta_{1j} F_{ji})}, \tag{6.3}$$

where $Y_i = 1$ for cryptocurrencies, $Y_i = 0$ for classical assets, and $F_j, j \in \{1, \ldots, k\}$ are the k orthogonal factors retrieved through the Factor Analysis. Based on the explanatory power and the significance of model 6.3, we can derive the most important factors contributing to the specific difference of cryptocurrencies. As a performance measure for Model 6.3, we are using $\tilde{R^2}$ (Nagelkerke, 1991), where:

$$\tilde{R}^2 = \frac{1 - \left\{\frac{L(\mathbf{0})}{L(\widehat{\boldsymbol{\beta}})}\right\}^{\frac{2}{n}}}{1 - \{L(\mathbf{0})\}^{\frac{2}{n}}}. \tag{6.4}$$

In Equation 6.4, $L(0)$ is the likelihood of the intercept-only model, $L(\widehat{\boldsymbol{\beta}})$ is the likelihood of the full model, and $\widehat{\boldsymbol{\beta}}$ is the vector of Maximum Likelihood estimated parameters.

Discriminant Analysis

The aim of discriminant analysis is to classify one or more observations into *a priori* known groups, minimizing the error of misclassification (Härdle and Simar, 2012). Formally, Linear Discriminant Analysis (LDA) assumes that the input dataset is multivariate Normal: $X_i \sim \mathrm{N}(\mu_i, \Sigma)$, where X_i belong to class ω_i. The

goal is to project samples X onto a line $Z = w^\top X$, where we select the projection that maximizes the standardized separability of the means over all directions. Specifically, we maximize the normalized, squared distance in the means of the classes

$$w^* = \arg\max_w \frac{|w^\top(\mu_i - \mu_j)|^2}{s_i^2 + s_j^2}, \tag{6.5}$$

$$s_i^2 = \sum_{x_i \in \omega_i} (w^\top x_i - w^\top \mu_i)^2 = w^\top S_i w, \tag{6.6}$$

giving the Linear Discriminant of Fisher (1936):

$$w^* = S_W^{-1}(\mu_i - \mu_j), \; S_W = S_i + S_j. \tag{6.7}$$

Quadratic Discriminant Analysis (QDA) follows the same procedure, but for $X_i \sim \mathrm{N}(\mu_i, \Sigma_i)$ belong to the class ω_i, one can relax the condition of equality of covariance matrices by $\Sigma_i \neq \Sigma_j$, $i \neq j$, allowing for a non-linear classifier.

Support Vector Machines

Support Vector Machines (SVM) is a data classification technique, its goal being to produce a model which predicts target values based on a set of attributes (Cristianini and Shawe-Taylor, 2000). The goal is to find a projection that maximizes margin in a hyperplane of the original data, without any parametric assumptions on the underlying stochastic process. The support vectors are determined via a quadratic optimization problem i.e. given a training data set D with n samples and 2 dimensions $D = (X_1, Y_1), \ldots (X_n, Y_n)$, $X_i \in \mathbb{R}^2$, $\quad Y_i \in [0, 1]$, the aim is to find a hyperplane that maximizes the margin:

$$\min_{w,b} \frac{1}{2}\|w\|^2, \text{ s.t. } \quad Y_i \left(w^\top X_i + b\right) \geq 1, i = 1, \ldots, n. \tag{6.8}$$

K-means Clustering Algorithm

This clustering method was first popularized by (MacQueen, 1967), who acknowledge a couple of other researchers that independently used that method around the same time. The aim is to allocate each observation of a data set in one of $k \in \mathbb{N}$ clusters, where k is predefined, so as to minimize the within-cluster sums of squares. In brief, the algorithm proceeds as follows:

i. Take k data points and set them as the cluster centres.

ii. Iteratively, for each data point, assign it to the cluster which centre is closer to the data point (the Euclidean distance is usually used, but other distance metrics have been proposed). Update the cluster centre for the selected cluster.

iii. Repeat until convergence (*i.e.* the allocations do not change).

Variance Components Split methods: MVCS, GMVCS

These methods aim to separate, respectively, the components of a structure like the types of assets herein or the types of Iris flowers, and clusters defined as the components of a mixture distribution. They are based on an unusual variance decomposition in between-group variations (Yatracos, 1998, 2013). To describe the sample version of the decomposition, let $X_1, \ldots, X_n$ be i.i.d. random variables. $X_{(j)}$ is the j-th order statistic, $1 \leq j \leq n$.
Consider the groups $X_{(1)}, \ldots, X_{(i)}$ and $X_{(i+1)}, \ldots, X_{(n)}$ with averages, respectively, $\bar{X}_{[1,i]}$ and $\bar{X}_{[i+1,n]}$, $i = 1, ..., n-1$, then

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n-1}\frac{i(n-i)}{n^2}(\bar{X}_{[i+1,n]} - \bar{X}_{[1,i]})(X_{(i+1)} - X_{(i)}). \tag{6.9}$$

The summands on the right side of Equation 6.9 measure between-groups variations. The standardized sample variance components

$$W_i = W_i(X_1, \ldots, X_n) \tag{6.10}$$

$$= \frac{i(n-i)}{n} \frac{(\bar{X}_{[i+1,n]} - \bar{X}_{[1,i]})(X_{(i+1)} - X_{(i)})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \; i = 1, \ldots, n-1, \tag{6.11}$$

indicate the relative contribution of the groups $X_{(1)}, ..., X_{(i)}$ and $X_{(i+1)}, ..., X_{(n)}$ in the sample variability. The index

$$\mathcal{I}_n = \max\{W_i, i = 1, \ldots, n-1\} \tag{6.12}$$

determines two potential clusters or parts of a structure and is based on averages and inter-point distances. When $\mathcal{I}_n = W_j$, these clusters are $\tilde{\mathcal{C}}_\infty = \{\mathcal{X}_{(\infty)}, \ldots, \mathcal{X}_{(|)}\}$, $\tilde{\mathcal{C}}_\in = \{\mathcal{X}_{(|+\infty)}, \ldots, \mathcal{X}_{(\backslash)}\}$. The observed $\mathcal{I}_n$-value is significant at α-level for the normal model when it exceeds the critical value $[-\ln(-\ln(1-\alpha)) + \ln n]/n$ (Yatracos, 2009); $\alpha = 0.05$ is used herein.

When $\mathcal{X}$ is the n by r data matrix of r-dimensional observations, $\mathbf{X}_j$ is the j-th row of $\mathcal{X}, j = 1, \ldots, n$. The coefficients of the orthogonal projection of $\mathcal{X}$ along the unit norm r-row vector $\mathbf{a}$ are $\mathcal{X}\mathbf{a} = (\mathbf{X}_1\mathbf{a}, \ldots, \mathbf{X}_n\mathbf{a})$.

The split in the sorted values of $\mathcal{X}\mathbf{a}$, where

$$\mathcal{I}_\mathcal{X}(\mathbf{a}) = \max\{W_i(\mathbf{X}_1\mathbf{a}, \ldots, \mathbf{X}_n\mathbf{a}); i = 1, ..., n-1\} \tag{6.13}$$

is attained, determines *along* $\mathbf{a}$ the groups $\tilde{\mathcal{C}}_{\mathcal{X},\infty}(\mathbf{a})$ and $\tilde{\mathcal{C}}_{\mathcal{X},\in}(\mathbf{a})$ in the $\mathcal{X}$-rows which are potential clusters and parts of a structure. For example, if for the data herein $\tilde{\mathcal{C}}_{\mathcal{X},\infty}(\mathbf{a})$ consists of rows 1-14, cryptocurrencies (a component) among the assets (the structure) are completely separated along $\mathbf{a}$.

The Maximum Variance Component Split (MVCS) method compares known components of a structure, *e.g.* cryptocurrencies herein, with data splits for a set of unit projection directions $\mathcal{D}_M$ usually determined by M positive equidistant angles

of $[0, \pi]$; *e.g.* when $r = 2$ and $M = 3$ the angles used are $\pi/3, 2\pi/3, \pi$. When one of the data split along projection direction $\mathbf{a}$ coincides with a component of the structure we have complete separation of this component along $\mathbf{a}$.

A set of projection directions $\mathcal{D}_M$ can be

$$(\Pi_{l=1}^{r} cos\theta_l,\ sin\theta_1 \Pi_{l=2}^{r} cos\theta_l, ...,\ sin\theta_{r-1} cos\theta_r,\ sin\theta_r), \tag{6.14}$$

where θ_l takes values in $\{\frac{m\pi}{M}, m = 1, ..., M\}, l = 1, ..., r$.

The method is computationally intensive for large r and M values, thus it may be used on subsets of the $\mathcal{X}$-columns. The importance of a subset S of $\mathcal{X}$-columns in the separation of a structure's component is measured by the number N_S of projection directions 6.14 completely separating the component. Indications for the importance of a specific column c in S in the separation of the same component are obtained by comparing N_S with the number of projection directions N_{S-c} separating the component when c is left out and also by comparing all $N_{S-c}, c \in S$. Similar indications of importance can be used for subgroups of S-columns.

The Global Maximum Variance Component Split (GMVCS) along all projection vectors $\mathcal{D}$, to be obtained from $\max\{\mathcal{I}_{\mathcal{X}}(\mathbf{a}), \mathbf{a} \in \mathcal{D}\}$ that is called the index determines two clusters. In practice, its approximation is obtained using $\mathcal{D}_M$. The splitting of these clusters may continue (Yatracos, 2013).

Expanding window modelling

For observing the assets dynamic, we are using an expanding window approach, allowing to distinguish the evolution of the clusters. In fact, for $t \in \{t_0, .., T\}$, where $t_0 = [T/3]$, the p-dimensional dataset X_t is projected on the k-dimensional space defined by the main factors extracted through the Factor Analysis applied on the dataset X_T. By using this projection instead of a time-varying factor model, we are avoiding situations like changes in factors loadings, causing inconsistencies over time.

6.3 Data and Results

Our dataset is a combination of cryptocurrencies and classical assets (commodities, exchange rates and stocks), covering the time period 01/02/2014 - 08/30/2019 (1426 trading days), for $n = 679$ assets (see Table 6.1). The reason for choosing this time span for the analysis is that before 2015 the liquidity in the cryptocurrency market had been relatively low, their total market capitalization being less than US$16 billion (Feng et al., 2018).

Table 6.1: Dataset

Type of Asset	Number of Assets	Source
Cryptocurrencies	150	Coinmarketcap
Stocks	496	Bloomberg
Exchange rates	13	Bloomberg
Commodities	20	Bloomberg

For robustness purposes, only the assets with at least 500 observations were kept in the analysis.[1] The first component of the dataset contains a representative sample of 150 cryptocurrencies selected from the top 500 cryptocurrencies sourced from https://coinmarketcap.com/, accounting for 98% of total market capitalization. The second component contains a sample of the most traded commodities indexes, the third component contains a sample of the most liquid exchange rates, while the fourth component contains the constituents of the S&P500 Index, recorded at August 30th 2019. As cryptocurrencies daily data are available at all times, while the stocks data obtained from Bloomberg observe market closure days (weekends

[1]The complete list of the assets included in the analysis can be found here.

and public holidays), the cryptocurrency data were pre-processed and the returns were computed in the same way as for classical assets (for example, the Monday return compares the Monday closing price to the Friday closing price).

6.3.1 Factor Analysis

Factor Analysis is a classical method used to find latent variables or factors among observed variables, by grouping variables with similar characteristics. Three steps are involved:

i. Estimation of the correlation matrix for all the variables, shown in Figure 6.1.

ii. Extraction of the factors from the correlation matrix, based on the correlation coefficients of the variables.

iii. Factor rotation, in order to maximize the relationship between the variables and relevant factors.

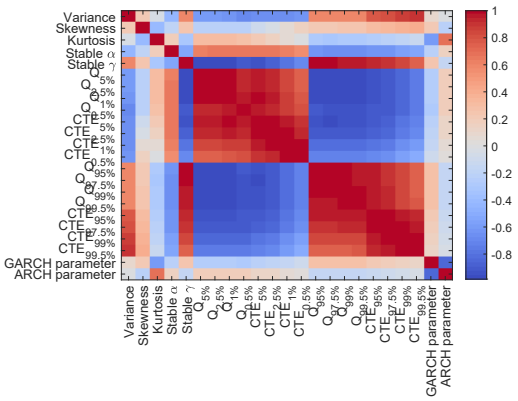

Figure 6.1: Correlation matrix

SFA_Cryptos

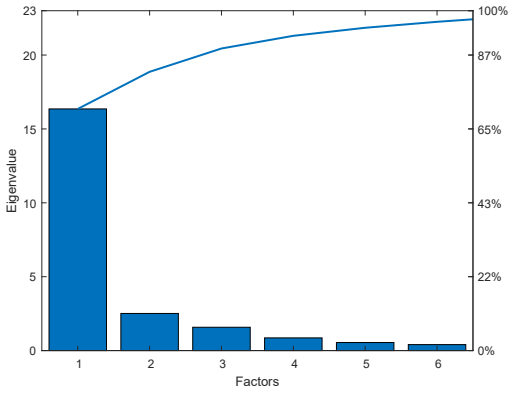

Figure 6.2: Scree plot

SFA_Cryptos

Based on the eigenvalues criteria, we can select those factors for which the eigenvalue is higher than 1 (see Figure 6.2). Accordingly, three factors were selected,

accounting for 89% of the total variance. In order to test the sampling adequacy of the Factor Analysis, we are using the Kaiser-Meyer-Olkin (KMO) test, which should be greater than 0.5 for a satisfactory Factor Analysis (Tabachnick and Fidell, 2013). The overall KMO test is computed as:

$$KMO = \frac{\sum_i \sum_{i \neq j} r_{ij}^2}{\sum_i \sum_{i \neq j} r_{ij}^2 + \sum_i \sum_{i \neq j} u_{ij}^2}. \tag{6.15}$$

where $R = (r_{ij})_{\substack{i=1\dots n \\ j=1\dots n}}$ is the correlation matrix and $U = (u_{ij})_{\substack{i=1\dots n \\ j=1\dots n}}$ is the partial covariance matrix (Cerny and Kaiser, 1977; Kaiser, 1974).

In fact, the KMO measure represents the proportion of the variance in the input variables that might be caused by underlying factors (Kaiser, 1981). In our sample, the overall KMO value is 0.92, pointing out that the Factor Analysis is suitable for structure detection. For the factor rotation, we used the VARIMAX method, which outputs orthogonal factors, also minimizing the number of variables that have high loadings on each factor. Based on the rotated factors pattern, the following conclusions can be drawn (see Figure 6.4):

i. First factor: **the tail factor**, accounting for 71% of the total variance, is highly correlated with the following parameters: the tail parameter alpha and the scale parameter gamma of the stable distribution, the lower and upper quantiles of the distribution of log-returns, the conditional tail expectations and the variance of log-returns.

ii. Second factor: **the moment factor**, accounting for 11% of the total variance, is highly correlated with the variance and the skewness coefficient of the distribution of log-returns.

iii. Third factor: **the memory factor**, accounting for 7% of the total variance, is highly correlated with the GARCH and ARCH parameters of the

GARCH(1,1) model estimated for log-returns and also the kurtosis of the distribution of log-returns.

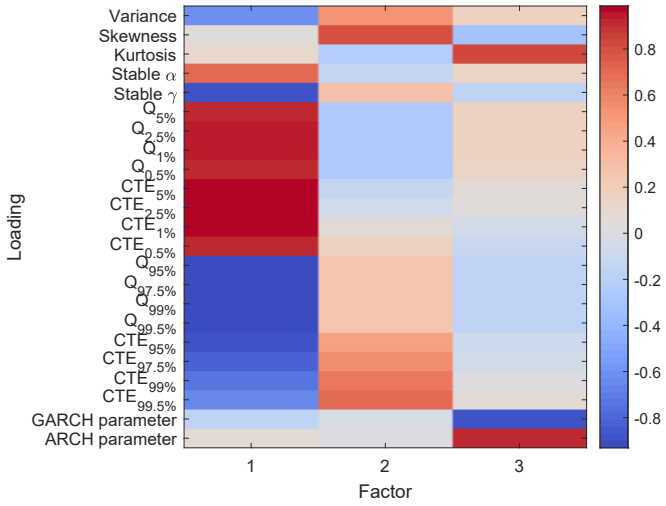

Figure 6.3: Loadings of the three factors

SFA_Cryptos

Based on the data revealed in Table 6.2, one can synthesize few characteristics of cryptocurrencies that differentiate them from the other assets. First, cryptocurrencies have higher variance of the log-return's distribution, compared to classical assets. Second, cryptocurrencies have longer tails, as indicated by the high values of quantiles and conditional tail expectations, i.e. cryptocurrencies have higher propensity for extreme values, in both tails of the log-returns distribution.

Based on the factors estimated through the Factor Analysis, one can map cryptocurrencies and classical assets, in order to derive some clustering effect.

Figures 6.4 and 6.5 map cryptocurrencies and classical assets; the colour code is the following: green: cryptocurrencies, black: stocks, red: commodities, blue: exchange rates. Also, a 95% confidence region is estimated, based on the Bivariate Kernel Density; in all figures only the top 10 cryptocurrencies (according to their market capitalization) are labeled.

Table 6.2: Assets profile based on the average values of the initial variables

Variable	**Commodities**	**Cryptocurrencies**	**Exchange rates**	**Stocks**
$\sigma^2 \cdot 10^3$	3.603	43.274	0.027	1.260
$Skewness$	0.214	3.876	−1.231	−7.797
$Stable_\alpha$	1.713	1.398	1.703	1.692
$Stable_\gamma \cdot 10^3$	9.266	47.080	2.868	8.738
$Q_{0.5\%}$	−0.026	−0.159	−0.008	−0.025
$Q_{1\%}$	−0.034	−0.211	−0.010	−0.033
$Q_{2.5\%}$	−0.043	−0.300	−0.012	−0.045
$Q_{5\%}$	−0.054	−0.388	−0.014	−0.056
$CTE_{0.5\%}$	−0.042	−0.274	−0.011	−0.047
$CTE_{1\%}$	−0.056	−0.367	−0.013	−0.065
$CTE_{2.5\%}$	−0.082	−0.546	−0.017	−0.108
$CTE_{5\%}$	−0.122	−0.744	−0.020	−0.167
$CTE_{95\%}$	0.044	0.368	0.011	0.038
$CTE_{97.5\%}$	0.058	0.533	0.013	0.049
$CTE_{99\%}$	0.087	0.877	0.015	0.072
$CTE_{99.5\%}$	0.128	1.299	0.018	0.099
$Q_{95\%}$	0.026	0.171	0.007	0.024
$Q_{97.5\%}$	0.034	0.246	0.010	0.030
$Q_{99\%}$	0.046	0.377	0.012	0.040
$Q_{99.5\%}$	0.057	0.518	0.014	0.050
$ARCH$	0.111	0.494	0.079	0.698
$GARCH$	0.665	0.478	0.720	0.206
$Kurtosis$	58.608	218.732	38.167	561.702

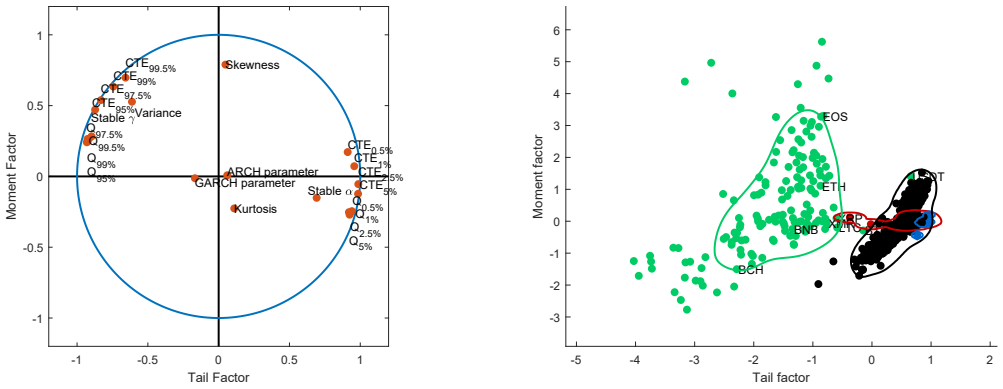

Figure 6.4: Loadings (left) and scores (right) based on tail and moment factor

SFA_cryptos

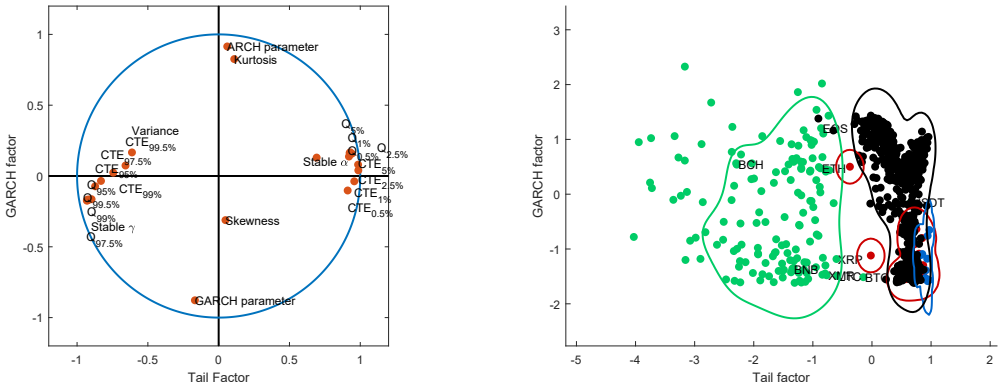

Figure 6.5: Loadings (left) and scores (right) based on tail and memory factor

SFA_Cryptos

As shown in Figures 6.4 and 6.5, there is a clear separation between cryptocurrencies and classical assets, mainly due to the first factor, the tail factor, while the memory and moment factor are of subliminal importance. The projection on three dimensional space defined by the Factor Analysis reveals two cryptocurrencies with atypical behaviour: Bitcoin and Tether. Thus, Bitcoin (BTC), the oldest and the most traded cryptocurrency, is closer to classical stocks and commodities, in terms of the tail factor, i.e. Bitcoin can be considered at the border between

classical assets and cryptocurrencies. Tether (USDT), on the other hand, a token that attempts to be tied to the US dollar, is indistinguishable from classical assets.

6.3.2 Assets classification

In this section, we list the results of the models presented in Section 2.3, in order to assess the ability of the factors produced through the Factor Analysis to discriminate between cryptocurrencies and classical assets. First, for each of the three factors we estimated the Binary Logistic model

$$P(Y_i = 1) = \frac{\exp(\beta_{0j} + \beta_{1j} F_{ji})}{1 + \exp(\beta_{0k} + \beta_{1j} F_{ji})}, \tag{6.16}$$

where $Y_i = 1$ for cryptocurrencies, $Y_i = 0$ for classical assets, and $F_j, j \in \{1, 2, 3\}$ are the orthogonal factors retrieved through the Factor Analysis. Table 6.3 lists the estimated β_{1j} of the Binary Logistic Regression model 6.16, with the performance measure defined by Equation 6.4.

Table 6.3: Estimates of Binary logistic regression model

Exogenous factor	**Factor 1**	**Factor 2**	**Factor 3**
Estimated β_1	-7.879**	0.728**	-0.389**
	(1.077)	(0.102)	(0.093)
$\tilde{R^2}$	0.967	0.134	0.034

Note: Standard errors in parentheses; ** denotes significance at 95% confidence level.

As seen in Figure 6.3, the most important factor regarding the separation between cryptocurrencies and classical assets is the tail factor, while the other two factors have little influence. Second, we employed Discriminant Analysis and Support Vec-

tor Machines on the space defined by the two first factors (tail and moment). Figure 6.6 illustrates the classification results using Discriminant Analysis. Quadratic classifiers have a good classification power, the only cryptocurrencies which are misclassified being Bitcoin and Tether.

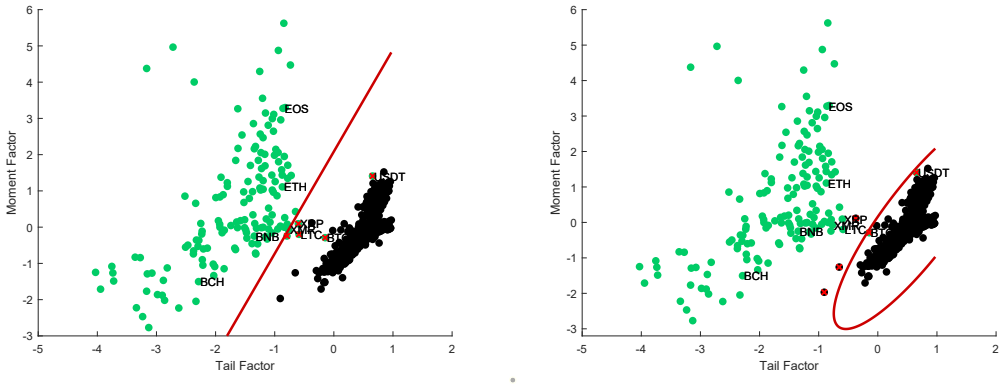

Figure 6.6: Discriminant Analysis: linear (left) and quadratic (right). Green dots denote cryptocurrencies, while the black dots denote the other assets; the dots highlighted in red are cases of misclassification

SFA_Cryptos

The same conclusion can be drawn by looking at the results of the Support Vector Machines non-linear classifier, according to which all cryptocurrencies are correctly classified using the tail factor and the moment factor (see Figure 6.7).

The k-means clustering algorithm (MacQueen, 1967) was also used for $k = 2, 3, 4$, the results showing that this method does not provide perfect classification[2]. The optimal number of clusters, as determined by the Elbow method, is $k = 10$; however, five clusters contain only cryptocurrencies, while the the other five clusters contain stocks, commodities and exchange rates, plus cryptocurrencies Bitcoin and Tether. In order to control for the influence of classical assets, these five clusters containing stocks, commodities, exchange rates, Bitcoin and Tether were merged

[2]Perfect classification is the case when a specific component is completely separated by the rest. In other words, all the members of that component, and only those, are in one cluster.

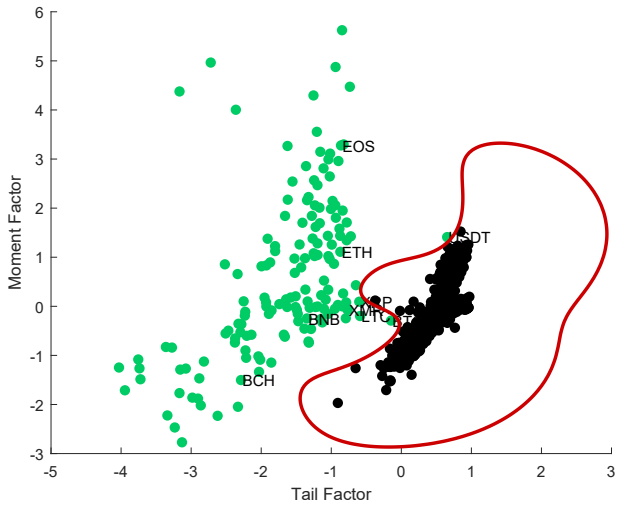

Figure 6.7: Support Vector Machines classification with green dots denoting cryptocurrencies, while black dots denote the other asset classes

SFA_Cryptos

into a single cluster. The final six clusters are shown in Figure 6.8, projected onto three dimensional space defined by the three factors extracted trough the Factor Analysis; the black dots denote classical assets (including Bitcoin and Tether), while the colored dotes denote cryptocurrencies. Each cryptocurrencies cluster was labeled with its leader in terms of market capitalization.

Thus, cryptocurrencies clusters are the following:

- Bitcoin and Tether cluster, having a similar behavior with classical assets,
- Ripple (XRP) cluster,
- Bitcoin Cash (BCH) cluster,
- Komodo (KMD) cluster,
- Ethereum (ETH) cluster,
- EOS cluster.

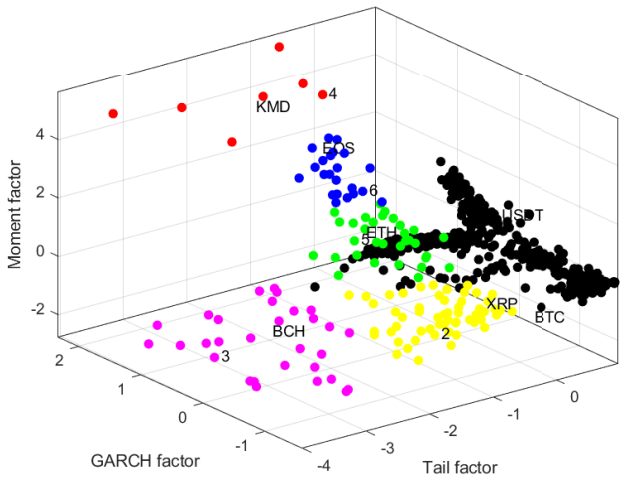

Figure 6.8: Projection of the clusters on the 3D space extracted trough Factor Analysis

Cluster_Cryptos

As shown in Table 6.4, there is a large variability in clusters profiles: the KMD cluster is the most volatile, with the highest variance and longest tails. The XRP cluster and BTC&USDT have the highest memory parameter, indicating a stronger persistence of volatility shocks. The validity of this classification is proven by the Multiclass Error-Correcting Output Codes (ECOC) model using Support Vector Machines (SVM) binary learners (Allwein et al., 2000), obtaining a classification accuracy of 99.41%, based on the three orthogonal factors extracted trough Factor Analysis (see Figure 6.9).

The results when applying the MVCS method, where the goal is to achieve separation of the components of the assets, are in accordance with those of Binary Logistic Regression, Discriminant Analysis, K-means clustering and Support Vector Machines. In order to apply MVCS method, we are considering the following term structure: the assets-data are regarded as a matrix $X_T = (x_{iT,k})_{\substack{i=1\dots n \\ k=1\dots p}} \in M(n,p)$, with $p = 23$ columns, representing the variables used for taxonomy, and $n = 679$, representing the assets. To be concise, the 23 columns are considered to be

Table 6.4: Crytptocurrencies clusters profile

Variable	**BTC&USDT**	**XRP**	**BCH**	**KMD**	**ETH**	**EOS**
σ^2	0.01	0.02	0.05	0.23	0.03	0.07
$Skewness$	16.13	1.61	−13.34	21.37	10.01	17.60
$Stable_\alpha$	1.05	1.39	1.36	1.41	1.46	1.45
$Stable_\gamma$	0.01	0.05	0.05	0.05	0.05	0.05
$Q_{0.5\%}$	−0.04	−0.16	−0.18	−0.16	−0.16	−0.16
$Q_{1\%}$	−0.06	−0.21	−0.24	−0.20	−0.20	−0.20
$Q_{2.5\%}$	−0.08	−0.30	−0.36	−0.33	−0.28	−0.27
$Q_{5\%}$	−0.10	−0.39	−0.46	−0.56	−0.34	−0.33
$CTE_{0.5\%}$	−0.08	−0.26	−0.39	−0.31	−0.23	−0.23
$CTE_{1\%}$	−0.09	−0.33	−0.57	−0.44	−0.29	−0.29
$CTE_{2.5\%}$	−0.13	−0.46	−1.00	−0.70	−0.38	−0.39
$CTE_{5\%}$	−0.18	−0.58	−1.55	−1.02	−0.45	−0.46
$CTE_{95\%}$	0.13	0.32	0.36	0.63	0.37	0.45
$CTE_{97.5\%}$	0.17	0.43	0.48	1.06	0.54	0.72
$CTE_{99\%}$	0.32	0.64	0.67	2.18	0.92	1.36
$CTE_{99.5\%}$	0.55	0.84	0.83	4.04	1.38	2.22
$Q_{95\%}$	0.04	0.17	0.20	0.16	0.16	0.16
$Q_{97.5\%}$	0.06	0.25	0.29	0.23	0.24	0.23
$Q_{99\%}$	0.08	0.38	0.43	0.36	0.36	0.37
$Q_{99.5\%}$	0.10	0.52	0.59	0.53	0.51	0.47
$ARCH$	0.23	0.17	0.61	0.70	0.66	0.89
$GARCH$	0.76	0.79	0.39	0.30	0.29	0.11
Kurtosis	551.59	43.33	348.62	514.30	210.30	384.02
Nr. Assets.	2	57	30	7	32	22

Note: column names denote the clusters leaders, according to their market capitalization.

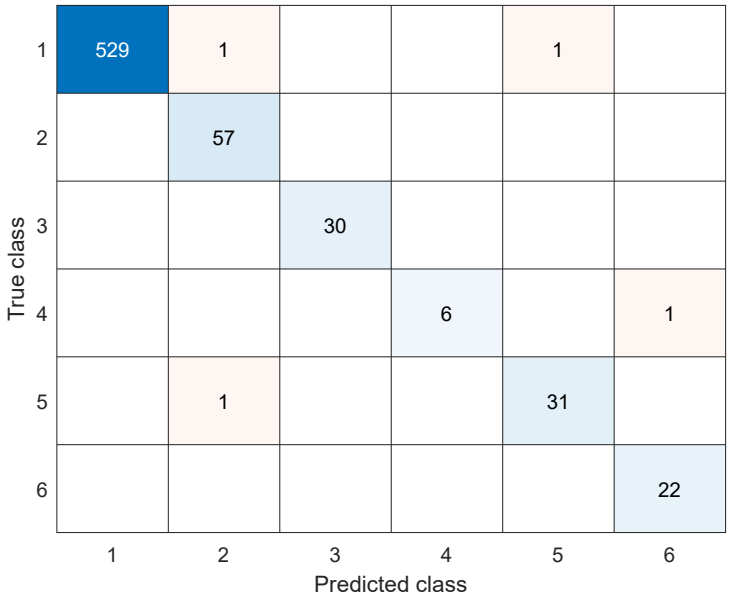

Figure 6.9: Confusion matrix of the ECOC classifier: 1 - classical assets, BTC and USDT, 2 - XRP cluster, 3 - BCH cluster, 4 - KMD cluster, 5 - ETH cluster, 6 - EOS cluster

Cluster_Cryptos

ordered, using the following order: $Variance$, $Skewness$, $Kurtosis$, $Stable_{\alpha}$, $Stable_{\gamma}$, $Q_{0.5\%}$, $Q_{1\%}$, $Q_{2.5\%}$, $Q_{5\%}$, $Q_{95.5\%}$, $Q_{97.5\%}$, $Q_{99\%}$, $Q_{99.5\%}$, $CTE_{0.5\%}$, $CTE_{1\%}$, $CTE_{2.5\%}$, $CTE_{5\%}$, $CTE_{95\%}$, $CTE_{97.5\%}$, $CTE_{99\%}$, $CTE_{99.5\%}$, $GARCH$, $ARCH$. The following notations are used for the MVCS method: M are the positive equidistant angles of $[0; \pi]$, S is a specific subset of the columns, N_S is the number of projection directions giving perfect classification when S is used, P_S is the corresponding percentage of these directions, while $minI, maxI$ are the minimum and the maximum index I value for perfect classification, respectively. The critical value for significance of the index for $\alpha = 5\%$ and $n = 679$ is 0.014.

In the following, we are presenting the results of the MVCS method for perfect classification of cryptocurrencies from the other assets, as it was found that for all three other structures (stocks, exchange rates and commodities), none of the

combinations of M and S presented below provided perfect classification.

First, due to processing power constraints, we split the data in two parts: the first part consists of columns 1-12 and the second part includes columns 13-23. For the same reason, projection directions (6.14) are used only for $M = 3, 6$. The number of projection directions used is M^{d-1}, with d, respectively, 12 and 11. Using all the data, perfect classification was not obtained for either data set. By omitting Tether (USDT), however, perfect classification was obtained for the fist set. Table 6.5 shows the results of the MVCS method, for all data, except USDT.

Table 6.5: Results of the MVCS method

M	S	N_S	P_S	$minI$	$maxI$
3	1-12	33	0.019%	0.041	0.135
6	1-12	27701	0.007%	0.045	0.146
3	13-23	0	0	n/a	n/a
6	13-23	0	0	n/a	n/a

The above results indicate that columns 1-12 are more important than columns 13-23. Table 6.6 shows the results of the MVCS method, when splitting these columns into disjoint subsets 1-6 and 7-12 (with USDT included). We can conclude that the most important columns for complete separation are columns 1-12, and in particular columns 1-6 (as can be seen from P_S, for the same value of M). The projected values for all the assets, using columns 1-6 and $M = 15$, on the projection direction that provided the largest index value among those that gave perfect classification of thre cryptocurrencies, are shown in Figure 6.10.

The projection direction that gave the largest index value for columns 1-6 and $M = 15$ is: $(0.005, -0.001, 0, 0.052, 0.497, 0.866)$. The index value in this direction is 0.048. In total 2 (out of 759375 tried) projection directions gave perfect

Table 6.6: Results of the MVCS method, columns 1-6 and 7-12

M	S	N_S	P_S	$minI$	$maxI$
3	1-6	0	0	n/a	n/a
6	1-6	0	0	n/a	n/a
9	1-6	0	0	n/a	n/a
12	1-6	0	0	n/a	n/a
15	1-6	2	0.0003%	0.038	0.048
18	1-6	0	0	n/a	n/a
3	7-12	0	0	n/a	n/a
6	7-12	0	0	n/a	n/a
9	7-12	0	0	n/a	n/a
12	7-12	0	0	n/a	n/a
15	7-12	0	0	n/a	n/a
18	7-12	0	0	n/a	n/a

classification and all provided statistically significant index values for the normal model.

Next, the first six columns are further used, as they are deemed the most important, according to the above. Then, the MVCS method is applied to all six quintets (these quintets are derived by omitting one of the six columns). Again, higher values of M are used., and the results being reported in Table 6.7.

One can see that the least important column is the third (Kurtosis), since its omission still provided perfect classification for the cryptocurrencies. As mentioned before, perfect classification is obtained only for cryptocurrencies, while the

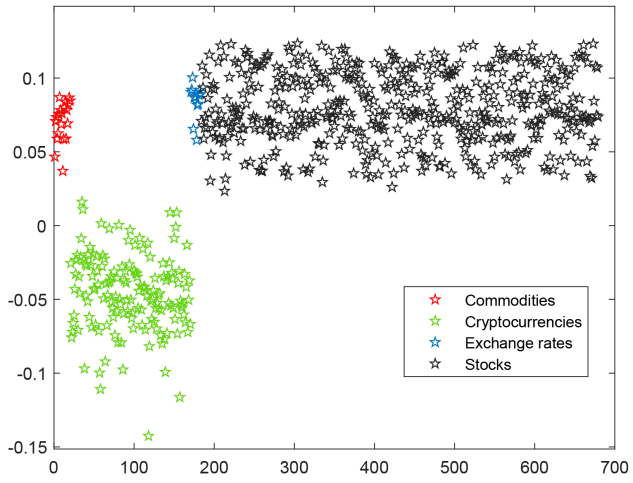

Figure 6.10: Projections of a subset of the data (the first 6 columns) for $M = 15$ on the projection direction that gave the largest index value among those that gave perfect classification of the cryptocurrencies

other assets have indistinct behaviour. This result is in line with the conclusions obtained trough the other classification techniques used above (Binary Logistic Regression, Discriminant Analysis and Support Vector Machines), MVCS method showing that cryptocurrencies behaves like a totally different species in the assets universe. Finally, it is worth noting that if USDT is excluded, by following the above procedure, we find that columns 7-12 are more important than columns 1-6. Among those columns, the least important seems to be the ninth ($Q_{0.005}$) and the most important seem to be the eighth ($Q_{0.01}$) and the eleventh ($CTE_{0.025}$). We also get more projection directions that provided perfect classification of the cryptocurrencies. This is in accordance to the observation that USDT is closely linked to the US dollar. We can conclude that cryptocurrencies are financial instruments whose specific difference is the tail behaviour of the distribution of daily log-returns. In other words, based on the tail factor profile, we can conclude that a random asset is likely to be a cryptocurrency if it has the following properties:

Table 6.7: Results for cryptocurrencies, all leave-one-out quintets of columns 1-6

M	S	N_S	P_S	$minI$	$maxI$
18	1,2,3,4,5	0	0	n/a	n/a
18	1,2,3,4,6	0	0	n/a	n/a
18	1,2,3,5,6	0	0	n/a	n/a
18	1,2,4,5,6	0	0	n/a	n/a
18	1,3,4,5,6	0	0	n/a	n/a
18	2,3,4,5,6	0	0	n/a	n/a
24	1,2,3,4,5	0	0	n/a	n/a
24	1,2,3,4,6	0	0	n/a	n/a
24	1,2,3,5,6	0	0	n/a	n/a
24	1,2,4,5,6	0	0	n/a	n/a
24	1,3,4,5,6	0	0	n/a	n/a
24	2,3,4,5,6	0	0	n/a	n/a
32	1,2,3,4,5	0	0	n/a	n/a
32	1,2,3,4,6	0	0	n/a	n/a
32	1,2,3,5,6	0	0	n/a	n/a
32	1,2,4,5,6	61	0.006%	0.034	0.077
32	1,3,4,5,6	0	0	n/a	n/a
32	2,3,4,5,6	0	0	n/a	n/a

very long tails of the log-returns distribution (in terms of the left and right quantile and the conditional tail expectation), high variance, high value of the α-stable scale parameter and value of the α-stable tail index close to 1.

6.4 Synchronic evolution of cryptocurrencies

For observing the assets dynamic, we are using an expanding window approach, allowing to distinguish the evolution of the clusters. In fact, for $t = t_0, \ldots, T$, the p-dimensional dataset is projected on the k-dimensional space defined by the main factors extracted through the Factor Analysis applied on the dataset X_T. By using this projection instead of a time-varying factor model, we are avoiding situations like changes in factors loadings, causing inconsistencies over time.
In order to derive the dynamics of the assets' universe, we used an expanding window approach, described below:

- The 23-dimensional dataset is estimated for the time interval $[1, t_0]$ with 1 being 01/02/2014 and t_0 10/31/2016.
- Time window is extended on a daily basis, up to T representing 08/30/2019 and for each step in time, the dataset is projected on the 2-dimensional space defined by the tail factor and the moment factor, estimated for the entire time period.

Figure 6.11 presents a snapshot of the evolution of the assets universe using the expanding window approach. [3] Looking at the evolution of the assets universe, it appears that individual cryptocurrencies tend to develop over time similar characteristics (synchronic evolution) that make them fully distinguishable from classical assets.
In order to test this behaviour, we are using the Likelihood Ratio associated to model 6.3, estimated using the expanding window approach previously described. The Likelihood Ratio for this model can be defined as:

[3] The daily evolution of the assets universe, for the period 10/31/2016-08/30/2019, is depicted in the video *Crypto_movie*, attached to this paper as supplementary material.

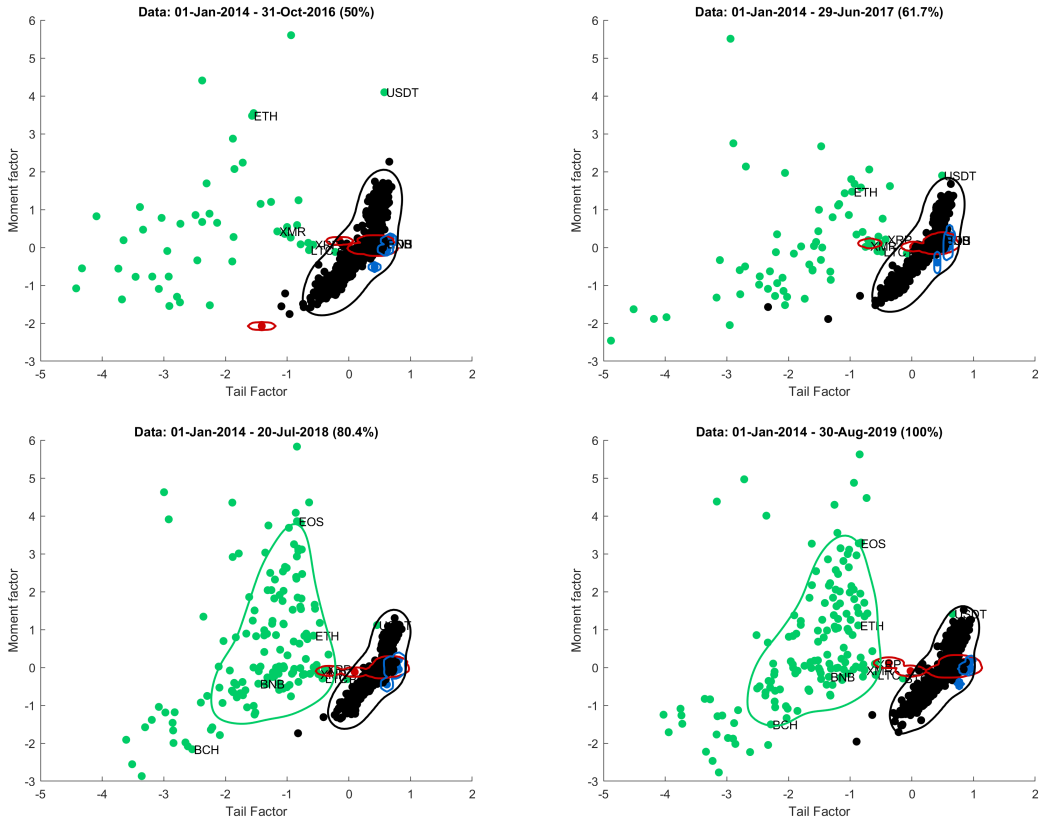

Figure 6.11: The evolution of the assets universe using the expanding window approach. The colour code is the following: green: cryptocurrencies, black: stocks, red: commodities, blue: exchange rates.

DFA_Cryptos

$$LR(\widehat{\boldsymbol{\beta}}) = -2(\log L(\widehat{\boldsymbol{\beta}}) - \log L(\widehat{\boldsymbol{\beta}_{\boldsymbol{s}}})), \tag{6.17}$$

where $L(\widehat{\boldsymbol{\beta}_{\boldsymbol{s}}})$ is the likelihood of a saturated model that fits perfectly the sample, while $L(\widehat{\boldsymbol{\beta}})$ is the likelihood of the estimated model. In the language of Binary Logistic Regression, the Likelihood Ratio from the Equation 6.17 is called deviance (Hosmer and Lemeshow, 2010) and is a measure of model goodness-of-fit, with large values indicating models with poor classification power. The deviance is

always positive, being zero only for the perfect fit. In order to derive the statistical significance of the classification, we compare the Likelihood Ratios of the estimated model and of the intercept-only model. Thus, we compute the difference of the likelihood ratios

$$D = LR(\widehat{\boldsymbol{\beta}}) - LR(0), \tag{6.18}$$

where asymptotically $D \sim \chi^2(1)$, $LR(0)$ being the likelihood ratio of the intercept-only model. In fact, we are estimating m models, where $m = T - t_0 - 1 = 740$. For each model we report the Likelihood Ratio (Figure 6.12) and the p-value associated to Equation 6.18 (see Figure 6.13). Large p-values indicates that the model might not differ statistically from an intercept-only model.

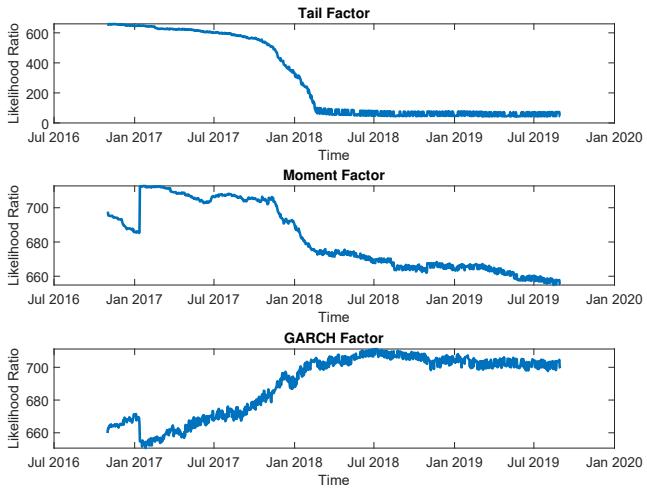

Figure 6.12: Likelihood Ratios for model (6.3), estimated on the time period 10/31/2016-08/30/2019, using an expanding window approach

CONV_Cryptos

By examining the evolution of the Likelihood Ratios, we can observe a trend change for the tail-factor-based model, starting January 2018, when the cryptocurrency

market collapsed after the historical maximum of Bitcoin from December 2017. Thus, the Likelihood Ratio converges to zero, pointing out the ability of the tail factor to discriminate between cryptocurrencies and classical assets.

The most important implication of this finding is the validity of synchronicity phenomenon among cryptocurrencies: in their evolution, the individual cryptocurrencies have developed similar characteristics (longer tails, higher volatility, higher propensity to extreme negative returns), that differentiate them from classical assets and position them as a new, different species in the ecosystem of financial instruments.

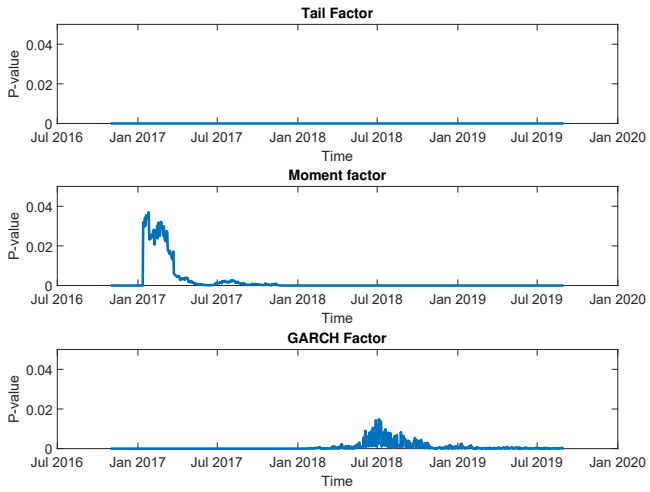

Figure 6.13: p-values for Equation 6.18, estimated on the time period 10/31/2016-08/30/2019, using an expanding window approach

CONV_Cryptos

6.5 Conclusions

In this paper we applied various classification techniques in order to discriminate between cryptocurrencies and classical assets, like stocks, exchange rates and com-

modities. Through the means of dimensionality reduction techniques and classification techniques, we proved that most of the variation among cryptocurrencies, stocks, exchanges rates and commodities can be explained by three factors: the tail factor, the moment factor and the memory factor. Our analysis revealed that the main difference between cryptocurrencies and classical assets, in terms of properties of the distribution of daily log-returns, is the tail behaviour, both in the left and in the right tail of the distribution. The moments of the distribution and the GARCH/ARCH parameters are of subliminal importance for discriminating between cryptocurrencies and classical assets.

Based on the tail factor profile, we can conclude that a random asset is likely to be a cryptocurrency if it has the following properties: very long tails of the log-returns distribution (in terms of the left and right quantile and the conditional tail expectation), high variance, high value of the α-stable scale parameter and value of the α-stable tail index closer to 1. Moreover, cryptocurrencies are completely separated by the other types of assets, as proved by Maximum Variance Components Split method. From the point of view of the risk analysts and regulators, the non-linear classification techniques applied on the factors extracted provide proficient results in order to discriminate between cryptocurrencies and the other assets.

Through the means of an expanding window approach, we are able to depict the evolutionary dynamics of cryptocurrencies universe and show how the clusters formed by projecting the multidimensional dataset on the main factors converge over time. By looking at the assets universe as a complex ecosystem, we are able to conclude that cryptocurrencies exhibit both a synchronic evolution (individual cryptocurrencies develop similar characteristics over time) and a divergent evolution, as different species, compared to classical assets.

Acknowledgements

Financial support from the Deutsche Forschungsgemeinschaft, Germany via IRTG 1792 “High Dimensional Non Stationary Time Series”, Humboldt-Universität zu Berlin, Germany, is gratefully acknowledged.

Appendix A: Cryptocurrency literature

Authors	Assets	Sample	Findings
Dyhrberg (2016a)	BTC, USD/EUR, USD/GBP, FTSE index.	2010-2015, daily data.	BTC can act as a hedge between UK equities and the USD.
Dyhrberg (2016b)	BTC, Federal funds rate, USD/EUR, USD/GBP, FTSE index, Gold futures, Gold cash.	2010-2015, daily data.	BTC is somewhere in between a currency (USD) and a commodity (Gold).
Bariviera et al. (2017)	BTC, USD/EUR, USD/GBP.	2011-2017, daily data. 2013-2016, hf data.	BTC presents large volatility and long-range memory (Hurst exponent higher than 0.5). BTC standard deviation is ten times greater than other currencies.
Baur et al. (2018)	BTC, Federal funds rate, USD/EUR, USD/GBP, FTSE index, Gold futures, Gold cash.	2010-2015, daily data.	BTC returns are not a hybrid of Gold and USD returns.
Caporale et al. (2018)	BTC, LTC, Ripple, Dash.	2013-2017, daily data.	The four cryptocurrencies exhibit persistence (Hurst exponent higher than 0.5), yet the persistency degree changes over time.
Härdle et al. (2018)	BTC, XRP, LTC, ETH, gold and S&P500.	2016-2018, daily data.	BTC, XRP, LTC, ETH exhibit higher volatility, skewness and kurtosis compared to Gold and S&P500 daily returns.
Henriques and Sadorsky (2018)	BTC and five exchange traded funds (ETFs): US equities (SPY), US bonds (TLT), US real estate (VNQ), Europe and Far East equities (EFA), and Gold (GLD).	2011-2017, daily data.	BTC can be a substitute for Gold in an investment portfolio, achieving a higher risk adjusted return.
Jiang et al. (2018)	BTC.	2010-2017, daily data.	Long-term memory and high degree of inefficiency ratio exists in the BTC market.
Klein et al. (2018)	BTC, CRIX index, Gold, Silver, crude oil, West Texas Intermediate (WTI), the S&P500, MSCI World and MSCI Emerging Markets 50.	2011-2017, daily data.	BTC returns have the highest mean and standard deviation.
Selmi et al. (2018)	BTC, Gold, Brent crude oil.	2011-2017, daily data.	Both BTC and Gold would serve the roles of a hedge, a safe haven and a diversifier for oil price movements.
Stosic et al. (2018)	Top 119 cryptocurrencies.	2016-2017, daily data.	Collective behaviour of the cryptocurrency market.
Takaishi (2018)	BTC, GBP/USD.	2014-2016, hf data.	The 1-min return distribution of BTC is fat-tailed, with high kurtosis; BTC time series exhibits multifractality.
Urquhart (2016)	BTC.	2010-2016, daily data.	Hurst statistic indicates strong anti-persistence (values lower than 0.5).
Wei (2018)	456 different cryptocurrencies.	2017, daily data.	Lower volatility for liquid cryptocurrencies. Illiquid cryptocurrencies exhibit strong return anti-persistence in the form of a low Hurst exponent.
Zhang et al. (2018)	70 % of cryptocurrencies market.	2013-2018, daily data.	Heavy tails, quickly decaying returns autocorrelations, slowly decaying autocorrelations for absolute returns, volatility clustering, leverage effects, long-range dependence, power-law correlation between price and volume.
Borri (2019)	BTC, ETH, LTC, XRP, Gold Bullion, the CBOE volatility index (VIX), the S&P400 and S&P500.	2017-2018, daily data.	Cryptocurrencies exhibit large and volatile return swings, and are riskier than most of the other assets.

Conclusion

Coming from a sphere in statistics and mathematics in which the Normal distribution is the dominating underlying stochastic term for the majority of the models, we indicate that the relevant diffusion, the Brownian Motion, although having appealing properties, is not accounting for three crucial empirical observations for financial data: Heavy tails, long memory and scaling laws.

Although there are many suitable applicants in the class of Lévy processes, which are able to account for heavy tails such as generalized hyperbolic (GH) distributions with the Student-t distribution as special case, the class of α-stable distributions is not only closed under convolution, but also obeys its own domain of attraction. This implies that in contrast to heavy-tailed, finite variance distributions, which follow the CLT and hence Normal behavior asymptotically, that the summation of α-stable random variables remains to be α-stable, non-normal for $\alpha < 2$ (GCLT). For multidimensional processes, elliptically stable distributions serve as generalization of the multidimensional Gaussian distribution. Although stock markets appear to be efficient in the sense that expected future price changes are difficult to forecast, functions of absolute price changes, which can be linked to second moments, are strongly auto-correlated. Although the majority of the literature focusses on short-memory processes, analyzing high-frequency data, we find strong evidence for long memory in higher sampling frequencies, starting at daily. A self-similar process, which is able to account for long-memory behavior is

the Fractional Brownian Motion, which has a possible non-Gaussian limit under convolution of the increments. The empirical scaling law states that the mean absolute (squared) price changes, as functions of their time intervals, are proportional to a power of the interval size. The increments of the Fractional Brownian Motion exhibit long memory through a parameter H, the Hurst exponent. For the Fractional Brownian Motion this scaling (Hurst) exponent would be constant over different orders of moments, being unifractal. But empirically, we observe varying Hölder exponents, the continuum of Hurst exponents, which implies multifractal behavior.

Centrally, we explain the multifractal behavior on the one hand through the changing α-stable indices over sampling frequencies by applying filters for seasonality and time dependence (long memory) over different sampling frequencies, starting at high-frequencies up to one minute. By accounting for intraday seasonalities, we obtain long memory behavior for sampling frequencies daily and higher. By utilizing a filter for long memory we show, that the low-sampling frequency process, not containing the time dependence component, can be governed by the α-stable motion. Under the α-stable motion we propose a semiparametric method coined Frequency Rescaling Methodology (FRM), which allows to rescale the filtered high-frequency data set to the lower sampling frequency, which is highly relevant for financial processes like stocks in order to estimate risk measures and allocate capital in portfolio allocation. The data sets for e.g. weekly data which we obtain by rescaling high-frequency data with the Frequency Rescaling Method (FRM) are more heavy tailed than we observe empirically. By employing an out-of-sample test, we show that using a subset of the whole data set (training set) suffices for the FRM to obtain a better forecast for the whole data set in terms of total variation distance. Specifically, the FRM would have been able to account for tail events of the financial crisis 2008 by increasing estimates for Value at Risk or Expected Shortfall. The same is expected to hold for the future: Utilizing high-frequency

events as a perspective of scaling up data points, which have never existed in the low-frequency, we incorporate future crises of possibly large magnitudes, providing higher forecasted risk, compared to existing low-frequency data. In contrast to modeling approaches involving universal function approximators, most often applied in machine learning, we resort to use an underlying parametric structure in form of self-similar processes in order to model the stock return distribution especially for lower sampling frequencies in a robust way.

The open question remains to be if the data series we observe are just a preasymptotic snapshot of a fractional Lévy stable motion, in which the Hölder exponents have converged. In that case the only relevant filter for the higher frequencies remains to be microstructure noise and seasonality, which are generated due to market makers and cyclical business components. If so, the diverging behavior within the sampling frequencies of the process is merely a relic of lacking data points in the lower sampling frequencies.

Bibliography

C. Acerbi. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26:1505 – 1518, 2002.

P. Algeot and T. Cover. Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *Annals of probability*, 16:876 – 898, 1988.

E. L. Allwein, R. E. Schapire, and Y. Singer. Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research*, 1:113 – 141, 2000.

T. Andersen and T. Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2-3):115 – 158, 1997.

T. Andersen and T. Bollerslev. Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance*, 53(1):219 – 265, 1998.

T. Andersen, T. Bollerslev, F. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579 – 625, 2003.

S. Bai, M. S. Taqqu, and T. Zhang. A unified approach to self-normalized block sampling. *Stochastic Processes and their Applications*, 126(8):2465 – 2493, 2016.

R. Baillie, T. Bollerslev, and H. O. Mikkelsen. Fractionally integrated generalized

autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3 – 30, 1996.

A. F. Bariviera, M. J. Basgall, W. Hasperué, and M. Naiouf. Some stylized facts of the Bitcoin market. *Physica A: Statistical Mechanics and its Applications*, 484:82 – 90, 2017.

D. J. Bartholomew. *Analysis of multivariate social science data*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. CRC Press, Boca Raton, Florida, 2011.

S. Basak and A. Shapiro. Value-at-risk-based risk management: Optimal policies and asset prices. *Review of Financial Studies*, 14(2):371 – 405, 2001.

D. G. Baur, T. Dimpfl, and K. Kuck. Bitcoin, gold and the US dollar - A replication and extension. *Finance Research Letters*, 25:103 – 110, 2018.

J. L. Bicksler and E. O. Thorp. The capital growth model: an empirical investigation. *Journal of Financial and Quantitative Analysis*, 8(2):273 – 287, 1973.

BIS. Supervisory framework for the use of backtesting in conjunction with the internal models approach to market risk capital requirements. Consultative Document, 1996.

BIS. Fundamental review of the trading book: A revised market risk framework. Consultative Document, 2013.

BIS. Minimum capital requirements for market risk. Consultative Document, 2016.

BIS. Basel iii: Finalising post-crisis reforms. Consultative Document, 2017.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307 – 327, 1986.

T. Bollerslev and I. Domowitz. Trading patterns and prices in the interbank foreign exchange market. *The Journal of Finance*, 48(4):1421 – 1443, 1993.

T. Bollerslev, J. Cai, and F. M. Song. Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the us treasury bond market. *Journal of Empirical Finance*, 7(1):37 – 55, 2000.

P. Boothe and D. Glassman. The statistical distribution of exchange rates. *Journal of International Economics*, 22:297 – 319, 1987.

N. Borri. Conditional tail-risk in cryptocurrency markets. *Journal of Empirical Finance*, 50:1 – 19, 2019.

L. Breiman. Optimal gambling system for favorable games. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, 1:63 – 68, 1961.

S. Browne. Survival and growth with a liability: Optimal portfolios in continuous time. *Math of Operations Research*, 22:468 – 493, 1997.

S. Browne. Risk-constrained dynamic active portfolio management. *Management Science*, 46(9):1188 – 1199, 2000.

S. Browne and W. Whitt. Portfolio choice and the bayesian kelly criterion. *Advances in Applied Probability*, 28(4):1145 – 1176, 1996.

E. Busseti, E. K. Ryu, and S. Boyd. Risk-constrained kelly gambling. *The Journal of Investing*, 25(3):118 – 134, 2016.

L. Calvet and A. Fisher. Multifractality in asset returns: Theory and evidence. *The Review of Economics and Statistics*, 84(3):381–406, 2002.

L. Calvet, A. J. Fisher, and B. Mandelbrot. Large deviations and the distribution of price changes. *Cowles Foundation Discussion Paper*, 1164, 1997.

G. M. Caporale, L. Gil-Alana, and A. Plastun. Persistence in the cryptocurrency market. *Research in International Business and Finance*, 46:141 – 148, 2018.

B. A. Cerny and H. F. Kaiser. A Study Of A Measure Of Sampling Adequacy For Factor-Analytic Correlation Matrices. *Multivariate behavioral research*, 12(1):43 – 47, 1977.

A. P. Chaboud, C. B., E. Hjalmarsson, and M. Loretan. Frequency of observation and the estimation of integrated volatility in deep and liquid financial markets. *Journal of Empirical Finance*, 17(2):212 – 240, 2010.

Y. Chen, W. Härdle, and V. Spokoiny. Ghica - risk analysis with gh distributions and independent components. *Journal of Empirical Finance*, 17(2):255–269, 2010.

V. Chopra and W. T. Ziemba. The effect of errors in the mean, variance, and covariance estimates on optimal portfolio choice. *Journal of Portfolio Management*, winter:6 – 11, 1993.

P. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39(4):841–62, 1998.

P. Cizek, W. Härdle, and R. Weron. *Statistical Tools for Finance and Insurance*. Springer, New York, 2011.

R. Clark and W. T. Ziemba. Playing the turn-of-the-year effect with index futures. *Operations Research*, 35(6):799 – 813, 1987.

A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661 – 703, 2009.

S. Corbet, B. Lucey, A. Urquhart, and L. Yarovaya. Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, 62: 182 – 199, 2019.

N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, Cambridge, 2000.

M. M. Dacorogna, R. Gencay, U. Müller, R. B. Olsen, and O. V. Pictet. *An introduction to high frequency finance*. Academic Press, New York, 2001.

J. Daníelsson and J.-P. Zigrand. On time-scaling of risk and the square-root-of-time rule. *Journal of Banking & Finance*, 30(10):2701 – 2713, 2006.

J. Davidson. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economic Statistics*, 22(1):16–29, 2004.

Z. Ding, C. W. J. Granger, and R. F. Engle. Long memory property of stock market returns and a new model. *Journal of Empirical Finance*, (1):83 – 106, 1993.

R. L. Dobrushin and P. Major. Non-central limit theorems for non-linear functional of gaussian fields. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 50(1):27 – 52, 1979.

A. H. Dyhrberg. Bitcoin, gold and the dollar - a garch volatility analysis. *Finance Research Letters*, 16:85 – 92, 2016a.

A. H. Dyhrberg. Hedging capabilities of bitcoin. Is it the virtual gold? *Finance Research Letters*, 16:139 – 144, 2016b.

A. ElBahrawy, L. Alessandretti, A. Kandler, R. Pastor-Satorras, and A. Baronchelli. Evolutionary dynamics of the cryptocurrency market. *Royal Society Open Science*, 4(11):170623, 2017.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987 – 1007, 1982.

R. F. Engle and E. Sokalska. Forecasting intraday volatility in the us equity market: Multiplicative component garch. *Journal of Financial Econometrics*, 10(1):54 – 83, 2012.

E. F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1): 34 – 105, 1965.

E. F. Fama. *Foundations of Finance: Portfolio Decisions and Securities Prices*. Basic Books, New York, 1976.

E. F. Fama and R. Roll. Properties of symmetric stable distributions. *Journal of the American Statistical Association*, 63(323):817 – 836, 1968.

W. Feng, Y. Wang, and Z. Zhang. Can cryptocurrencies be a safe haven: a tail risk perspective analysis. *Applied Economics*, 50(44):4745 – 4762, 2018.

T. Fischer, C. Krauss, and A. Deinert. Statistical arbitrage in Cryptocurrency markets. *Journal of Risk and Financial Management*, 12(1):31, 2019.

A. J. Fisher, L. Calvet, and B. Mandelbrot. Multifractality of Deutschemark/US dollar exchange rates. *Cowles Foundation Discussion Paper*, 1166, 1997.

R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179 – 188, 1936.

I. M. Fund. Treatment of Crypto assets in macroeconomic statistics. *International Monetary Fund, Statistics Department*, 2019.

P. Giot. Market risk models for intraday data. *The European Journal of Finance*, 11(4):415 – 428, 2005.

B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley Mathematics Series, Cambridge, MA, 1954a.

B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley Mathematics Series, Cambridge, MA, 1954b.

M. Grabchak and G. Samorodnitsky. Do financial returns have finite or infinite variance? A paradox and an explanation. *Quantitative Finance*, 10(8):883–893, 2010.

N. H. Hakansson. Optimal investment and consumption strategies under risk for a class of utility functions. *Econometrica*, 38:587 – 607, 1970.

N. H. Hakansson. On optimal myopic portfolio policies, with and without serial correlation of yields. *Journal of Business*, 44:324 – 334, 1971.

B. E. Hansen. Shrinkage efficiency bounds. *Econometric Theory*, 31:860 – 879, 2015.

W. Härdle and L. Simar. *Applied multivariate statistical analysis: [R & Matlab codes]*. Springer, Berlin and Heidelberg, 3. ed. edition, 2012.

W. K. Härdle, C. Harvey, and R. Reule. Understanding Cryptocurrencies. *IRTG 1792 Discussion Paper*, 2018-044, 2018.

P. Hartmann, S. Straetmans, and C. G. de Vries. Heavy tails and currency crises. *Journal of Empirical Finance*, 17(2):241 – 254, 2010.

D. B. Hausch and W. T. Ziemba. Transactions costs, extent of inefficiencies, entries and multiple wagers in a racetrack betting model. *Management Science*, 31(4): 381 – 394, 1985.

I. Henriques and P. Sadorsky. Can Bitcoin replace Gold in an investment portfolio? *Journal of Risk and Financial Management*, 11(3):48, 2018.

B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163 – 1174, 1975.

D. W. Hosmer and S. Lemeshow. *Applied logistic regression.* A Wiley-Interscience publication. John Wiley, New York, 2010.

W. James and C. Stein. Estimation with quadratic loss. *Proc. Third Berkeley Symp. Math. Statist. Prob.*, 1:361 – 379, 1961.

Y. Jiang, H. Nie, and W. Ruan. Time-varying long-term memory in Bitcoin market. *Finance Research Letters*, 25:280 – 284, 2018.

H. F. Kaiser. An index of factorial simplicity. *Psychometrika*, 39(1):31 – 36, 1974.

H. F. Kaiser. A revised measure of sampling adequacy for Factor-Analytic data matrices. *Educational and Psychological Measurement*, 41(2):379 – 381, 1981.

P. Kearns and A. Pagan. Estimating the density tail index for financial time series. *The Review of Economics and Statistics*, 79(2):171 – 175, 1997.

J. Kelly. A new interpretation of information rate. *Bell System Technology Journal*, 35:917 – 926, 1956.

T. Klein, H. Pham Thu, and T. Walther. Bitcoin is not the New Gold - A comparison of volatility, correlation, and portfolio performance. *International Review of Financial Analysis*, 59:105 – 116, 2018.

A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell' Istituto Italianodegli Attuari*, 4:83 – 91, 1933.

I. A. Koutrouvelis. Regression-Type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75(372):918, 1980.

I. A. Koutrouvelis. An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics - Simulation and Computation*, 10 (1):17 – 28, 1981.

S. Kring, S. Rachev, M. Höchstötter, and F. Fabozzi. Estimation of alpha-stable sub-gaussian distributions for assets returns. *Risk Assessment: Decisions in Banking and Finance*, pages 111 – 152, 2009.

P. Kupiec. Techniques for verifying the accuracy of risk measurement models. Finance and Economics Discussion Series 95-24, Board of Governors of the Federal Reserve System (U.S.), 1995.

P. Lévy. *Calcul des probabilités*. Gauthier-Villars, Paris, 1925.

J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1): 13 – 37, 1965.

Y. Lv and B. K. Meister. Appliation of the kelly criterion to ornstein-uhlenbeck processes. *Lecture notes of the institute for Computer Sciences*, 4:1051 – 1062, 2009.

L. MacLean and W. T. Ziemba. Capital growth: Theory and practice. *Working Paper*, 2005.

L. C. MacLean, W. T. Ziemba, and G. Blazenko. Growth versus security in dynamic investment analysis. *Management Science*, 38(11):1562 – 1585, 1992.

L. C. MacLean, W. T. Ziemba, and Y. Li. Time to wealth goals in capital accumulation. *Quantitative Finance*, 5(4):343 – 355, 2005.

L. C. MacLean, E. O. Thorp, and W. T. Ziemba. *The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific Press, Singapore, 2011.

J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281 – 297, Berkeley, Calif, 1967. University of California Press.

B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36:394 – 419, 1963a.

B. Mandelbrot. New methods in statistical economics. *Journal of Political Economy*, 71:421 – 440, 1963b.

B. Mandelbrot. The variation of some other speculative prices. *The Journal of Business*, 40:393 – 413, 1967.

B. Mandelbrot. *The Fractal Geometry of Nature*. Henry Holt and Company, New York, 1982.

B. Mandelbrot, A. J. Fisher, and L. Calvet. A multifractal model of asset returns. *Cowles Foundation Discussion Paper*, 1164, 1997.

B. Mandelbrot, P. H. Cootner, R. E. Gomory, E. F. Fama, W. S. Morris, and H. M. Taylor. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. Selecta Volume E*. Springer, New York, 1997b.

B. B. Mandelbrot and J. W. van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10:422 – 437, 1968.

R. N. Mantegna and H. E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376:46 – 49, 1995.

H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7:77 – 91, 1952.

H. M. Markowitz. Investment for the long run: New evidence for an old rule. *The Journal of Finance*, 31(5):1273 – 1286, 1976.

T. D. Matteo, T. Aste, and M. Dacorogna. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking and Finance*, 29(4):827–851, 2005.

J. H. McCulloch. Measuring tail thickness to estimate the stable index α: A critique. *Journal of Business and Economic Statistics*, 15(1):74 – 81, 1997.

J. W. McFarland, R. Pettit, and S. K. Sung. The distribution of foreign exchange price changes: Trading day effects and risk measurement. *Journal of Finance*, 37(3):693 – 715, 1982.

R. C. Merton. Lifetime portfolio selection under uncertainty: The continuous time case. *Review of Economics and Statistics*, 51:247 – 259, 1969.

R. C. Merton. An analytic derivation of the efficient portfolio frontier. *The Journal of Financial and Quantitative Analysis*, 7(4):1851 – 1872, 1972.

R. C. Merton. *Continuous Time Finance*. MA Blackwell Publishers Inc., Malden, 1992.

M. Mirtaheri, S. Abu-El-Haija, F. Morstatter, G. V. Steeg, and A. Galstyan. Identifying and analyzing cryptocurrency manipulations in social media. 2019.

U. A. Müller, M. Dacorogna, R. Olsen, O. V. Pictet, M. Schwarz, and C. Morgenegg. Statistical study of foreign exchange rates, empirical evidence of a price

change scaling law, and intraday analysis. *Journal of Banking and Finance*, 14 (6):1189 – 1208, 1990.

N. J. Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691 – 692, 1991.

J. P. Nolan. *Levy Processes: Maximum Likelihood Estimation and Diagnostics for Stable Distributions*. Birkhauser, Boston, 2001.

J. P. Nolan. Multivariate elliptically contoured stable distributions: theory and estimation. *Computational Statistics*, 28(5):2067 – 2089, 2013.

J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston (To be published), 2017.

J. A. Núñez, M. I. Contreras-Valdez, and C. A. Franco-Ruiz. Statistical analysis of Bitcoin during explosive behavior periods. *PloS one*, 14(3), 2019.

R. Osorio. A prospect-theory approach to the kelly criterion for fat-tail portfolios: The case of the student t-distribution. *Working papers series*, 2008.

M. Paolella, S. Mittnik, and S. Rachev. Stationarity of stable power-GARCH processes. *Journal of Econometrics*, 106(1):97–107, 2002.

S. Poon. Intertemporal portfolio selection. Lecture note in continuous time finance. University of Manchester, 2010.

S. Rachev. *Handbook of Heavy Tailed Distributions in Finance: Handbooks in Finance*. Handbooks in Finance. Elsevier Science, North Holland, 2003.

S. Rachev and S. Mittnik. *Stable Paretian models in finance*. Series in financial economics and quantitative analysis. J. Wiley, Chichester, 2000a.

S. Rachev and S. Mittnik. *Stable Paretian models in finance*. John Wiley and Sons, New York, 2000b.

P. Robinson. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47(1):67 – 84, 1991.

R. Roll. Evidence on the growth optimum model. *The Journal of Finance*, 28(3): 551 – 566, 1973.

D. Ruppert. *Statistics and Data Analysis for Financial Engineering*. Springer Texts in Statistics. Springer New York, 2010.

G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York, 1994.

R. Selmi, W. Mensi, S. Hammoudeh, and J. Bouoiyour. Is Bitcoin a hedge, a safe haven or a diversifier for oil price movements? A comparison with gold. *Energy Economics*, 74:787 – 801, 2018.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425 – 442, 1964.

N. V. Smirnov. Sui la distribution de w^2 (criterium de mr.v. mises). *Comptes Rendus (Paris)*, 202:449 – 452, 1936.

C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.*, 1:197 – 206, 1955.

D. L. Stevens. Financial characteristics of merged firms: A multivariate analysis. *The Journal of Financial and Quantitative Analysis*, 8(2):149, 1973.

D. Stosic, T. B. Ludermir, and T. Stosic. Collective behavior of cryptocurrency price changes. *Physica A: Statistical Mechanics and its Applications*, 507:499 – 509, 2018.

B. G. Tabachnick and L. S. Fidell. *Using multivariate statistics*. Pearson Education, Boston, 2013.

T. Takaishi. Statistical properties and multifractality of Bitcoin. *Physica A: Statistical Mechanics and its Applications*, 506:507 – 519, 2018.

M. S. Taqqu. Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 50(1):53 – 83, 1979.

S. J. Taylor. *Modelling Financial Time Series*. John Wiley & Sons, New York, 1986.

E. O. Thorp. Portfolio choice and the kelly criterion. *Proceedings of the Business and Economics Section of the American Statistical Association*, pages 215 – 224, 1971.

E. O. Thorp. *The Mathematics of Gambling*. John Wiley and Sons, New York, 1984.

E. O. Thorp. The kelly criterion in blackjack, sports betting and the stock market. *In S. A. Zenios and W. T. Ziemba (Eds.), Handbook of Asset and Liability Management*, 1:387 – 428, 2006.

J. Tobin. Liquidity preference as behavior towards risk. *The Review of Economic Studies*, 25(67):65 – 86, 1958.

Transparency Market Research. Cryptocurrency Market to reach US$ 6,702.1 mn by 2025. 2018.

A. Urquhart. The inefficiency of Bitcoin. *Economics Letters*, 148:80 – 82, 2016.

M. Veillette. STBL: Alpha stable distributions for MATLAB. *Matlab Central File Exchange*, 2012.

W. C. Wei. Liquidity and market efficiency in cryptocurrencies. *Economics Letters*, 168:21 – 24, 2018.

N. Wesselhöfft and W. K. Härdle. Risk-constrained Kelly portfolios under alpha-stable laws. *Computational Economics*, 2019.

R. Westerfield. The distribution of common stock prices changes: An application of transaction time and subordinated stochastic models. *The Journal of Financial and Quantitative Analysis*, 12(5):743 – 765, 1977.

Z. Xu and R. Gencay. Scaling, self-similarity and multifractality in FX markets. *Physica A: Statistical Mechanics and its Applications*, 323(C):578–590, 2003.

Y. G. Yatracos. Variance and clustering. *Proceedings of the American Mathematical Society*, 126(04):1177 – 1179, 1998.

Y. G. Yatracos. The asymptotic distribution of a cluster-index for i.i.d. normal random variables. *The Annals of Applied Probability*, 19(2):585 – 595, 2009.

Y. G. Yatracos. Detecting clusters in the data from variance decompositions of its projections. *Journal of Classification*, 30(1):30 – 55, 2013.

N. Yoshino and F. Taghizadeh-Hesary. Analytical framework on credit risks for financing small and medium-sized enterprises in Asia. *Asia-Pacific Development Journal*, 21(2):1 – 21, 2015.

W. Zhang, P. Wang, X. Li, and D. Shen. Some stylized facts of the cryptocurrency market. *Applied Economics*, 50(55):5950 – 5965, 2018.

V. M. Zolotarev. *One-dimensional Stable Distributions*, volume 65. American Mathematical Society, translations of mathematical monographs edition, 1986.

G. O. Zumbach, F. Corsi, and A. Trapletti. Efficient estimation of volatility using high frequency data. *Working Paper*, 2002.